P[illegible]ol[illegible]ical Bulletin

H[illegible]son, Editor
[illegible] of Texas

CONTENTS

Conv[illegible] Validation by the Multitrait-Multi-[illegible] . . . Don[illegible] T. Campbell and Donald W. Fiske 81

The [illegible] of Time [illegible] of Conditioned Reflex Activity . . . . A. [illegible] Dmitriev and A. M. Kochigina 106

The [illegible] Intens[illegible] Intermittent Sound on Performance, [illegible]logy . . . . Robert Plutchik 133

Method[illegible] Considerations [illegible] the Construct Validation of Drive-[illegible] . . . . Donald H. Kausler and E. Philip Trapp 152

On the [illegible] of Pr[illegible] Techniques . . . Gardner Lindzey 158

Published [illegible]ly by the
American [illegible]ical Association

Vol. [illegible] March, 1959

[illegible] *Psychological Bulletin* contains evaluative reviews of [illegible] and [illegible] on research methodology in psychology. This Journal [illegible] *original research or original theoretical articles.*

[illegible] *[illegible]uscripts* should be sent to the Editor, Harry Helson, [illegible] University of Texas, Austin, Texas.

[illegible]aration of articles for publication. Authors are strongly [illegible] the [illegible] directions given in the *Publication Manual of the Am[illegible]* [illegible] (1957 Revision). Special attention should be given [illegible] on [illegible] preparation of the references (pp. 50–60), since this is a [illegible] of [illegible] in long reviews of research literature. *All copy* [illegible], [illegible]ing *the references.* All manuscripts should be submitted [illegible] one of [illegible] should be an original typed copy; author's name should app[illegible] this [illegible]. Original figures are prepared for publication; duplicate figures [illegible] photo[illegible] or pencil-drawn copies. Authors are cautioned to retain a [illegible] to guard against loss in the mail and to check carefully the [illegible] final [illegible].

[illegible]prints. Fifty free offprints are given to contributors of articles [illegible]. Au[illegible] of early publication articles receive no gratis offprints.

[illegible] C. [illegible], Managing Editor; Helen Orr, Promotion Manager; [illegible]

[illegible]munications—including subscriptions, orders of back issues, [illegible] of [illegible]—should be addressed to the American Psychological Asso[illegible], 1333 Six[illegible] Street N.W., Washington 6, D. C. Address changes must re[illegible] Subscrip[illegible] Office by the 10th of the month to take effect the following m[illegible] [illegible] resulting from address changes will not be replaced; subscri[illegible] [illegible] office that they will guarantee second-class forwarding post[illegible]. [illegible] delivered copies must be made within four months of publica[illegible].

[illegible]nual subscription: $8.00 (Foreign $8.50). Single copies, $1.5[illegible]

PUBLISHED BIMONTHLY BY:
THE AMERICAN PSYCHOLOGICAL ASSOCIA[illegible]
Menasha, Wisconsin
and 1333 Sixteenth Street N.W., Washington 6, D.C.
[illegible] as second class matter at Washington, D.C., and Menasha, Wisconsin. [illegible] Washington, D.C., and at additional mailing offices. Printed in U.S.A.

Vol. 56, No. 2 March, 1959

Psychological Bulletin

CONVERGENT AND DISCRIMINANT VALIDATION BY THE MULTITRAIT-MULTIMETHOD MATRIX[1]

DONALD T. CAMPBELL
Northwestern University
AND DONALD W. FISKE
University of Chicago

In the cumulative experience with measures of individual differences over the past 50 years, tests have been accepted as valid or discarded as invalid by research experiences of many sorts. The criteria suggested in this paper are all to be found in such cumulative evaluations, as well as in the recent discussions of validity. These criteria are clarified and implemented when considered jointly in the context of a multitrait-multimethod matrix. Aspects of the validational process receiving particular emphasis are these:

1. Validation is typically *convergent*, a confirmation by independent measurement procedures. Independence of methods is a common denominator among the major types of validity (excepting content validity) insofar as they are to be distinguished from reliability.

2. For the justification of novel trait measures, for the validation of test interpretation, or for the establishment of construct validity, *discriminant* validation as well as convergent validation is required. Tests can be invalidated by too high correlations with other tests from which they were intended to differ.

3. Each test or task employed for measurement purposes is a *trait-method unit*, a union of a particular trait content with measurement procedures not specific to that content. The systematic variance among test scores can be due to responses to the measurement features as well as responses to the trait content.

4. In order to examine discriminant validity, and in order to estimate the relative contributions of trait and method variance, *more than one trait* as well as *more than one method* must be employed in the validation process. In many instances it will be convenient to achieve this through a multitrait-multimethod matrix. Such a matrix presents all of the intercorrelations resulting when each of several traits is measured by each of several methods.

To illustrate the suggested validational process, a synthetic example is

[1] The new data analyses reported in this paper were supported by funds from the Graduate School of Northwestern University and by the Department of Psychology of the University of Chicago. We are also indebted to numerous colleagues for their thoughtful criticisms and encouragement of an earlier draft of this paper, especially Benjamin S. Bloom, R. Darrell Bock, Desmond S. Cartwright, Loren J. Chapman, Lee J. Cronbach, Carl P. Duncan, Lyle V. Jones, Joe Kamiya, Wilbur L. Layton, Jane Loevinger, Paul E. Meehl, Marshall H. Segall, Thornton B. Roby, Robert C. Tryon, Michael Wertheimer, and Robert F. Winch.

TABLE 1

A SYNTHETIC MULTITRAIT-MULTIMETHOD MATRIX

		Method 1			Method 2			Method 3		
	Traits	A_1	B_1	C_1	A_2	B_2	C_2	A_3	B_3	C_3
Method 1	A_1	(.89)								
	B_1	.51	(.89)							
	C_1	.38	.37	(.76)						
Method 2	A_2	*.57*	.22	.09	(.93)					
	B_2	.22	*.57*	.10	.68	(.94)				
	C_2	.11	.11	*.46*	.59	.58	(.84)			
Method 3	A_3	*.56*	.22	.11	*.67*	.42	.33	(.94)		
	B_3	.23	*.58*	.12	.43	*.66*	.34	.67	(.92)	
	C_3	.11	.11	*.45*	.34	.32	*.58*	.58	.60	(.85)

Note.—The validity diagonals are the three sets of italicized values. The reliability diagonals are the three sets of values in parentheses. Each heterotrait-monomethod triangle is enclosed by a solid line. Each heterotrait-heteromethod triangle is enclosed by a broken line.

presented in Table 1. This illustration involves three different traits, each measured by three methods, generating nine separate variables. It will be convenient to have labels for various regions of the matrix, and such have been provided in Table 1. The reliabilities will be spoken of in terms of three *reliability diagonals*, one for each method. The reliabilities could also be designated as the monotrait-monomethod values. Adjacent to each reliability diagonal is the *heterotrait-monomethod* triangle. The reliability diagonal and the adjacent heterotrait-monomethod triangle make up a *monomethod block*. A *heteromethod block* is made up of a *validity* diagonal (which could also be designated as monotrait-heteromethod values) and the two *heterotrait-heteromethod* triangles lying on each side of it. Note that these two heterotrait-heteromethod triangles are not identical.

In terms of this diagram, four aspects bear upon the question of validity. In the first place, the entries in the validity diagonal should be significantly different from zero and sufficiently large to encourage further examination of validity. This requirement is evidence of convergent validity. Second, a validity diagonal value should be higher than the values lying in its column and row in the heterotrait-heteromethod triangles. That is, a validity value for a variable should be higher than the correlations obtained between that variable and any other variable having neither trait nor method in common. This requirement may seem so minimal and so obvious as to not need stating, yet an inspection of the literature shows that it is frequently not met,

and may not be met even when the validity coefficients are of substantial size. In Table 1, all of the validity values meet this requirement. A third common-sense desideratum is that a variable correlate higher with an independent effort to measure the same trait than with measures designed to get at different traits which happen to employ the same method. For a given variable, this involves comparing its values in the validity diagonals with its values in the heterotrait-monomethod triangles. For variables A_1, B_1, and C_1, this requirement is met to some degree. For the other variables, A_2, A_3 etc., it is not met and this is probably typical of the usual case in individual differences research, as will be discussed in what follows. A fourth desideratum is that the same pattern of trait interrelationship be shown in all of the heterotrait triangles of both the monomethod and heteromethod blocks. The hypothetical data in Table 1 meet this requirement to a very marked degree, in spite of the different general levels of correlation involved in the several heterotrait triangles. The last three criteria provide evidence for discriminant validity.

Before examining the multitrait-multimethod matrices available in the literature, some explication and justification of this complex of requirements seems in order.

Convergence of independent methods: the distinction between reliability and validity. Both reliability and validity concepts require that agreement between measures be demonstrated. A common denominator which most validity concepts share in contradistinction to reliability is that this agreement represent the convergence of independent approaches. The concept of independence is indicated by such phrases as "external variable," "criterion performance," "behavioral criterion" (American Psychological Association, 1954, pp. 13–15) used in connection with concurrent and predictive validity. For construct validity it has been stated thus: "Numerous successful predictions dealing with phenotypically diverse 'criteria' give greater weight to the claim of construct validity than do . . . predictions involving very similar behavior" (Cronbach & Meehl, 1955, p. 295). The importance of independence recurs in most discussions of proof. For example, Ayer, discussing a historian's belief about a past event, says "if these sources are numerous and independent, and if they agree with one another, he will be reasonably confident that their account of the matter is correct" (Ayer, 1954, p. 39). In discussing the manner in which abstract scientific concepts are tied to operations, Feigl speaks of their being "fixed" by "triangulation in logical space" (Feigl, 1958, p. 401).

Independence is, of course, a matter of degree, and in this sense, reliability and validity can be seen as regions on a continuum. (Cf. Thurstone, 1937, pp. 102–103.) Reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods. A split-half reliability is a little more like a validity coefficient than is an immediate test-retest reliability, for the items are not quite identical. A correlation between dissimilar subtests is probably a reliability measure, but is still closer to the region called validity.

Some evaluation of validity can take place even if the two methods

are not entirely independent. In Table 1, for example, it is possible that Methods 1 and 2 are not entirely independent. If underlying Traits A and B are entirely independent, then the .10 minimum correlation in the heterotrait-heteromethod triangles may reflect method covariance. What if the overlap of method variance were higher? All correlations in the heteromethod block would then be elevated, including the validity diagonal. The heteromethod block involving Methods 2 and 3 in Table 1 illustrates this. The degree of elevation of the validity diagonal above the heterotrait-heteromethod triangles remains comparable and relative validity can still be evaluated. The interpretation of the validity diagonal in an absolute fashion requires the fortunate coincidence of both an independence of traits and an independence of methods, represented by zero values in the heterotrait-heteromethod triangles. But zero values could also occur through a combination of negative correlation between traits and positive correlation between methods, or the reverse. In practice, perhaps all that can be hoped for is evidence for relative validity, that is, for common variance specific to a trait, above and beyond shared method variance.

Discriminant validation. While the usual reason for the judgment of invalidity is low correlations in the validity diagonal (e.g., the Downey Will-Temperament Test [Symonds, 1931, p. 337ff]) tests have also been invalidated because of too high correlations with other tests purporting to measure different things. The classic case of the social intelligence tests is a case in point. (See below and also [Strang, 1930; R. Thorndike, 1936].) Such invalidation occurs when values in the heterotrait-heteromethod triangles are as high as those in the validity diagonal, or even where within a monomethod block, the heterotrait values are as high as the reliabilities. Loevinger, Gleser, and DuBois (1953) have emphasized this requirement in the development of maximally discriminating subtests.

When a dimension of personality is hypothesized, when a construct is proposed, the proponent invariably has in mind distinctions between the new dimension and other constructs already in use. One cannot define without implying distinctions, and the verification of these distinctions is an important part of the validational process. In discussions of construct validity, it has been expressed in such terms as "from this point of view, a low correlation with athletic ability may be just as important and encouraging as a high correlation with reading comprehension" (APA, 1954, p. 17).

The test as a trait-method unit. In any given psychological measuring device, there are certain features or stimuli introduced specifically to represent the trait that it is intended to measure. There are other features which are characteristic of the method being employed, features which could also be present in efforts to measure other quite different traits. The test, or rating scale, or other device, almost inevitably elicits systematic variance in response due to both groups of features. To the extent that irrelevant method variance contributes to the scores obtained, these scores are invalid.

This source of invalidity was first noted in the "halo effects" found in ratings (Thorndike, 1920). Studies of individual differences among laboratory animals resulted in the recognition of "apparatus factors," usually more dominant than psychologi-

cal process factors (Tryon, 1942). For paper-and-pencil tests, methods variance has been noted under such terms as "test-form factors" (Vernon: 1957, 1958) and "response sets" (Cronbach: 1946, 1950; Lorge, 1937). Cronbach has stated the point particularly clearly: "The assumption is generally made . . . that what the test measures is determined by the content of the items. Yet the final score . . . is a composite of effects resulting from the content of the item and effects resulting from the form of the item used" (Cronbach, 1946, p. 475). "Response sets always lower the logical validity of a test. . . . Response sets interfere with inferences from test data" (p. 484).

While E. L. Thorndike (1920) was willing to allege the presence of halo effects by comparing the high obtained correlations with common sense notions of what they ought to be (e.g., it was unreasonable that a teacher's intelligence and voice quality should correlate .63) and while much of the evidence of response set variance is of the same order, the clear-cut demonstration of the presence of method variance requires both several traits and several methods. Otherwise, high correlations between tests might be explained as due either to basic trait similarity or to shared method variance. In the multitrait-multimethod matrix, the presence of method variance is indicated by the difference in level of correlation between the parallel values of the monomethod block and the heteromethod blocks, assuming comparable reliabilities among all tests. Thus the contribution of method variance in Test A_1 of Table 1 is indicated by the elevation of $r_{A_1B_1}$ above $r_{A_1B_2}$, i.e., the difference between .51 and .22, etc.

The distinction between trait and method is of course relative to the test constructor's intent. What is an unwanted response set for one tester may be a trait for another who wishes to measure acquiescence, willingness to take an extreme stand, or tendency to attribute socially desirable attributes to oneself (Cronbach: 1946, 1950; Edwards, 1957; Lorge, 1937).

Multitrait-Multimethod Matrices in the Literature

Multitrait-multimethod matrices are rare in the test and measurement literature. Most frequent are two types of fragment: two methods and one trait (single isolated values from the validity diagonal, perhaps accompanied by a reliability or two), and heterotrait-monomethod triangles. Either type of fragment is apt to disguise the inadequacy of our present measurement efforts, particularly in failing to call attention to the preponderant strength of methods variance. The evidence of test validity to be presented here is probably poorer than most psychologists would have expected.

One of the earliest matrices of this kind was provided by Kelley and Krey in 1934. Peer judgments by students provided one method, scores on a word-association test the other. Table 2 presents the data for the four most valid traits of the eight he employed. The picture is one of strong method factors, particularly among the peer ratings, and almost total invalidity. For only one of the eight measures, School Drive, is the value in the validity diagonal (.16!) higher than all of the heterotrait-heteromethod values. The absence of discriminant validity is further indicated by the tendency of the values in the monomethod triangles to approximate the reliabilities.

An early illustration from the ani-

TABLE 2

PERSONALITY TRAITS OF SCHOOL CHILDREN FROM KELLEY'S STUDY

($N=311$)

		Peer Ratings				*Association Test*			
		A_1	B_1	C_1	D_1	A_2	B_2	C_2	D_2
Peer Ratings									
Courtesy	A_1	(.82)							
Honesty	B_1	.74	(.80)						
Poise	C_1	.63	.65	(.74)					
School Drive	D_1	.76	.78	.65	(.89)				
Association Test									
Courtesy	A_2	*.13*	.14	.10	.14	(.28)			
Honesty	B_2	.06	*.12*	.16	.08	.27	(.38)		
Poise	C_2	.01	.08	*.10*	.02	.19	.37	(.42)	
School Drive	D_2	.12	.15	.14	*.16*	.27	.32	.18	(.36)

mal literature comes from Anderson's (1937) study of drives. Table 3 presents a sample of his data. Once again, the highest correlations are found among different constructs from the same method, showing the dominance of apparatus or method factors so typical of the whole field of individual differences. The validity diagonal for hunger is higher than the heteroconstruct-heteromethod values. The diagonal value for sex has not been *italicized* as a validity coefficient since the obstruction box measure was pre-sex-opportunity, the activity wheel post-opportunity. Note that the high general level of heterotrait-heteromethod values could be due either to correlation of methods variance between the two methods, or to correlated trait variance. On a priori grounds, however, the methods would seem about as independent as one would be likely to achieve. The predominance of an apparatus factor for the activity wheel is evident from the fact that the correlation between hunger and thirst

TABLE 3

MEASURES OF DRIVES FROM ANDERSON'S DATA

($N=50$)

		Obstruction Box			Activity Wheel		
		A_1	B_1	C_1	A_2	B_2	C_2
Obstruction Box							
Hunger	A_1	(.58)					
Thirst	B_1	.54	( )				
Sex	C_1	.46	.70	( )			
Activity Wheel							
Hunger	A_2	*.48*	.31	.37	(.83)		
Thirst	B_2	.35	*.33*	.43	.87	(.92)	
Post Sex	C_2	.31	.37	.44	.69	.78	( )

Note.—Empty parentheses appear in this and subsequent tables where no appropriate reliability estimates are reported in the original paper.

TABLE 4

SOCIAL INTELLIGENCE AND MENTAL ALERTNESS SUBTEST INTERCORRELATIONS FROM THORNDIKE'S DATA

($N=750$)

		Memory		Comprehension		Vocabulary	
		A_1	B_1	A_2	B_2	A_3	B_3
Memory							
Social Intelligence (Memory for Names & Faces)	A_1	( )					
Mental Alertness (Learning Ability)	B_1	.31	( )				
Comprehension							
Social Intelligence (Sense of Humor)	A_2	*.30*	.31	( )			
Mental Alertness (Comprehension)	B_2	.29	*.38*	.48	( )		
Vocabulary							
Social Intelligence (Recog. of Mental State)	A_3	*.23*	.35	*.31*	.35	( )	
Mental Alertness (Vocabulary)	B_3	.30	*.58*	.40	*.48*	.47	( )

(.87) is of the same magnitude as their test-retest reliabilities (.83 and .92 respectively).

R. L. Thorndike's study (1936) of the validity of the George Washington Social Intelligence Test is the classic instance of invalidation by high correlation between traits. It involved computing all of the intercorrelations among five subscales of the Social Intelligence Test and five subscales of the George Washington Mental Alertness Test. The model of the present paper would demand that each of the traits, social intelligence and mental alertness, be measured by at least two methods. While this full symmetry was not intended in the study, it can be so interpreted without too much distortion. For both traits, there were subtests employing acquisition of knowledge during the testing period (i.e., learning or memory), tests involving comprehension of prose passages, and tests that involved a definitional activity. Table 4 shows six of Thorndike's 10 variables arranged as a multitrait-multimethod matrix. If the three subtests of the Social Intelligence Test are viewed as three methods of measuring social intelligence, then their intercorrelations (.30, .23, and .31) represent validities that are not only lower than their corresponding monomethod values, but also lower than the heterotrait-heteromethod correlations, providing a picture which totally fails to establish social intelligence as a separate dimension. The Mental Alertness validity diagonals (.38, .58, and .48) equal or exceed the monomethod values in two out of three cases, and exceed all heterotrait-heteromethod control values. These results illustrate the general conclusions reached by Thorndike in his factor analysis of the whole 10×10 matrix.

The data of Table 4 could be used to validate specific forms of cognitive functioning, as measured by the different "methods" represented by usual intelligence test content on the one hand and social content on the other. Table 5 rearranges the 15 values for this purpose. The monomethod values and the validity diagonals exchange places, while the heterotrait-heteromethod control coefficients are the same in both tables.

TABLE 5

MEMORY, COMPREHENSION, AND VOCABULARY MEASURED WITH SOCIAL AND ABSTRACT CONTENT

		Social Content			Abstract Content		
		A_1	B_1	C_1	A_2	B_2	C_2
Social Content							
Memory (Memory for Names and Faces)	A_1	( )					
Comprehension (Sense of Humor)	B_1	.30	( )				
Vocabulary (Recognition of Mental State)	C_1	.23	.31	( )			
Abstract Content							
Memory (Learning Ability)	A_2	*.31*	.31	.35	( )		
Comprehension	B_2	.29	*.48*	.35	.38	( )	
Vocabulary	C_2	.30	.40	*.47*	.58	.48	( )

As judged against these latter values, comprehension (.48) and vocabulary (.47), but not memory (.31), show some specific validity. This transmutability of the validation matrix argues for the comparisons within the heteromethod block as the most generally relevant validation data, and illustrates the potential interchangeability of trait and method components.

Some of the correlations in Chi's (1937) prodigious study of halo effect in ratings are appropriate to a multitrait-multimethod matrix in which each rater might be regarded as representing a different method. While the published report does not make these available in detail because it employs averaged values, it is apparent from a comparison of his Tables IV and VIII that the ratings generally failed to meet the requirement that ratings of the same trait by different raters should correlate higher than ratings of different traits by the same rater. Validity is shown to the extent that of the correlations in the heteromethod block, those in the validity diagonal are higher than the average heteromethod-heterotrait values.

A conspicuously unsuccessful multitrait-multimethod matrix is provided by Campbell (1953, 1956) for rating of the leadership behavior of officers by themselves and by their subordinates. Only one of 11 variables (Recognition Behavior) met the requirement of providing a validity diagonal value higher than any of the heterotrait-heteromethod values, that validity being .29. For none of the variables were the validities higher than heterotrait-monomethod values.

A study of attitudes toward authority and nonauthority figures by Burwen and Campbell (1957) contains a complex multitrait-multimethod matrix, one symmetrical excerpt from which is shown in Table 6. Method variance was strong for most of the procedures in this study. Where validity was found, it was primarily at the level of validity diagonal values higher than heterotrait-heteromethod values. As illustrated in Table 6, attitude toward father showed this kind of validity, as did attitude toward peers to a lesser degree. Attitude toward boss showed no validity. There was no evidence of a generalized attitude toward authority which would include father and boss, although such values as the

TABLE 6

ATTITUDES TOWARD FATHER, BOSS, AND PEER, AS MEASURED BY INTERVIEW AND CHECK-LIST OF DESCRIPTIVE TRAITS

		Interview			Trait Check-List		
		A_1	B_1	C_1	A_2	B_2	C_2
Interview (N=57)							
Father	A_1	( )					
Boss	B_1	.64	( )				
Peer	C_1	.65	.76	( )			
Trait Check-List (N=155)							
Father	A_2	*.40*	.08	.09	(.24)		
Boss	B_2	.19	−*.10*	−.03	.23	(.34)	
Peer	C_2	.27	.11	*.23*	.21	.45	(.55)

.64 correlation between father and boss as measured by interview might have seemed to confirm the hypothesis had they been encountered in isolation.

Borgatta (1954) has provided a complex multimethod study from which can be extracted Table 7, illustrating the assessment of two traits by four different methods. For all measures but one, the highest correlation is the apparatus one, i.e., with the other trait measured by the same method rather than with the same trait measured by a different method. Neither of the traits finds

TABLE 7

MULTIPLE MEASUREMENT OF TWO SOCIOMETRIC TRAITS (N=125)

		Sociometric				Observation			
		by Others		by Self		Group Interaction		Role Playing	
		A_1	B_1	A_2	B_2	A_3	B_3	A_4	B_4
Sociometric by Others									
Popularity	A_1	( )							
Expansiveness	B_1	.47	( )						
Sociometric by Self									
Popularity	A_2	*.19*	.18	( )					
Expansiveness	B_2	.07	*.08*	.32	( )				
Observation of Group Interaction									
Popularity	A_3	*.25*	.18	*.26*	.11	( )			
Expansiveness	B_3	.21	*.12*	.28	*.15*	.84	( )		
Observation of Role Playing									
Popularity	A_4	*.24*	.14	*.18*	.01	*.66*	.58	( )	
Expansiveness	B_4	.25	*.12*	.26	*.05*	.66	*.76*	.73	( )

any consistent validation by the requirement that the validity diagonals exceed the heterotrait-heteromethod control values. As a most minimal requirement, it might be asked if the sum of the two values in the validity diagonal exceeds the sum of the two control values, providing a comparison in which differences in reliability or communality are roughly partialled out. This condition is achieved at the purely chance level of three times in the six tetrads. This matrix provides an interesting range of methodological independence. The two "Sociometric by Others" measures, while representing the judgments of the same set of fellow participants, come from distinct tasks: Popularity is based upon each participant's expression of his own friendship preferences, while Expansiveness is based upon each participant's guesses as to the other participant's choices, from which has been computed each participant's reputation for liking lots of other persons, i.e., being "expansive." In line with this considerable independence, the evidence for a method factor is relatively low in comparison with the observational procedures. Similarly, the two "Sociometric by Self" measures represent quite separate tasks, Popularity coming from his estimates of the choices he will receive from others, Expansiveness from the number of expressions of attraction to others which he makes on the sociometric task. In contrast, the measures of Popularity and Expansiveness from the observations of group interaction and the role playing not only involve the same specific observers, but in addition the observers rated the pair of variables as a part of the same rating task in each situation. The apparent degree of method variance within each of the two observational situations, and the apparent sharing of method variance between them, is correspondingly high.

In another paper by Borgatta (1955), 12 interaction process variables were measured by quantitative observation under two conditions, and by a projective test. In this test, the stimuli were pictures of groups, for which the *S* generated a series of verbal interchanges; these were then scored in Interaction Process Analysis categories. For illustrative purposes, Table 8 presents the five traits which had the highest mean communalities in the over-all factor analysis. Between the two highly similar observational methods, validation is excellent: trait variance runs higher than method variance; validity diagonals are in general higher than heterotrait values of both the heteromethod and monomethods blocks, most unexceptionably so for Gives Opinion and Gives Orientation. The pattern of correlation among the traits is also in general confirmed.

Of greater interest because of the greater independence of methods are the blocks involving the projective test. Here the validity picture is much poorer. Gives Orientation comes off best, its projective test validity values of .35 and .33 being bested by only three monomethod values and by no heterotrait-heteromethod values within the projective blocks. All of the other validities are exceeded by some heterotrait-heteromethod value.

The projective test specialist may object to the implicit expectations of a one-to-one correspondence between projected action and overt action. Such expectations should not be attributed to Borgatta, and are not necessary to the method here proposed. For the simple symmetrical model of this paper, it has been as-

TABLE 8

INTERACTION PROCESS VARIABLES IN OBSERVED FREE BEHAVIOR, OBSERVED ROLE PLAYING AND A PROJECTIVE TEST

($N=125$)

		Free Behavior					Role Playing					Projective Test				
		A_1	B_1	C_1	D_1	E_1	A_2	B_2	C_2	D_2	E_2	A_3	B_3	C_3	D_3	E_3
Free Behavior																
Shows solidarity	A_1	( )														
Gives suggestion	B_1	.25	( )													
Gives opinion	C_1	.13	.24	( )												
Gives orientation	D_1	−.14	.26	.52	( )											
Shows disagreement	E_1	.34	.41	.27	.02	( )										
Role Playing																
Shows solidarity	A_2	*.43*	.43	.08	.10	.29	( )									
Gives suggestion	B_2	.16	*.32*	.00	.24	.07	.37	( )								
Gives opinion	C_2	.15	.27	*.60*	.38	.12	.01	.10	( )							
Gives orientation	D_2	−.12	.24	.44	*.74*	.08	.04	.18	.40	( )						
Shows disagreement	E_2	.51	.36	.14	−.12	*.50*	.39	.27	.23	−.11	( )					
Projective Test																
Shows solidarity	A_3	*.20*	.17	.16	.12	.08	*.17*	.12	.30	.17	.22	( )				
Gives suggestion	B_3	.05	*.21*	.05	.08	.13	.10	*.19*	−.02	.06	.30	.32	( )			
Gives opinion	C_3	.31	.30	*.13*	−.02	.26	.25	.19	*.15*	−.04	.53	.31	.63	( )		
Gives orientation	D_3	−.01	.09	.30	*.35*	−.05	.03	.00	.19	*.33*	.00	.37	.29	.32	( )	
Shows disagreement	E_3	.13	.18	.10	.14	*.19*	.22	.28	.02	.04	*.23*	.27	.51	.47	.30	( )

TABLE 9

MAYO'S INTERCORRELATIONS BETWEEN OBJECTIVE AND RATING MEASURES OF INTELLIGENCE AND EFFORT ($N = 166$)

		Peer Ratings		Objective	
		A_1	B_1	A_2	B_2
Peer Rating					
Intelligence	A_1	(.85)			
Effort	B_1	.66	(.84)		
Objective Measures					
Intelligence	A_2	*.46*	.29	( )	
Effort	B_2	.46	*.40*	.10	( )

sumed that the measures are labeled in correspondence with the correlations expected, i.e., in correspondence with the traits that the tests are alleged to diagnose. Note that in Table 8, Gives Opinion is the best projective test predictor of both free behavior and role playing Shows Disagreement. Were a proper theoretical rationale available, these values might be regarded as validities.

Mayo (1956) has made an analysis of test scores and ratings of effort and intelligence, to estimate the contribution of halo (a kind of methods variance) to ratings. As Table 9 shows, the validity picture is ambiguous. The method factor or halo effect for ratings is considerable although the correlation between the two ratings (.66) is well below their reliabilities (.84 and .85). The objective measures share no appreciable apparatus overlap because they were independent operations. In spite of Mayo's argument that the ratings have some valid trait variance, the .46 heterotrait-heteromethod value seriously depreciates the otherwise impressive .46 and .40 validity values.

Cronbach (1949, p. 277) and Vernon (1957, 1958) have both discussed the multitrait-multimethod matrix shown in Table 10, based upon data originally presented by H. S. Conrad. Using an approximative technique, Vernon estimates that 61% of the systematic variance is due to a general factor, that 21½% is due to the test-form factors specific to verbal or to pictorial forms of items, and that but 11½% is due to the content fac-

TABLE 10

MECHANICAL AND ELECTRICAL FACTS MEASURED BY VERBAL AND PICTORIAL ITEMS

		Verbal Items		Pictorial Items	
		A_1	B_1	A_2	B_2
Verbal Items					
Mechanical Facts	A_1	(.89)			
Electrical Facts	B_1	.63	(.71)		
Pictorial Items					
Mechanical Facts	A_2	*.61*	.45	(.82)	
Electrical Facts	B_2	.49	*.51*	.64	(.67)

tors specific to electrical or to mechanical contents. Note that for the purposes of estimating validity, the interpretation of the general factor, which he estimates from the .49 and .45 heterotrait-heteromethod values, is equivocal. It could represent desired competence variance, representing components common to both electrical and mechanical skills—perhaps resulting from general industrial shop experience, common ability components, overlapping learning situations, and the like. On the other hand, this general factor could represent overlapping method factors, and be due to the presence in both tests of multiple choice item format, IBM answer sheets, or the heterogeneity of the *S*s in conscientiousness, test-taking motivation, and test-taking sophistication. Until methods that are still more different and traits that are still more independent are introduced into the validation matrix, this general factor remains uninterpretable. From this standpoint it can be seen that $21\frac{1}{2}\%$ is a very minimal estimate of the total test-form variance in the tests, as it represents only test-form components specific to the verbal or the pictorial items, i.e., test-form components which the two forms do *not* share. Similarly, and more hopefully, the $11\frac{1}{2}\%$ content variance is a very minimal estimate of the total true trait variance of the tests, representing only the true trait variance which electrical and mechanical knowledge do *not* share.

Carroll (1952) has provided data on the Guilford-Martin Inventory of Factors STDCR and related ratings which can be rearranged into the matrix of Table 11. (Variable R has been inverted to reduce the number of negative correlations.) Two of the methods, Self Ratings and Inventory scores, can be seen as sharing method variance, and thus as having an inflated validity diagonal. The more independent heteromethod blocks involving Peer Ratings show some evidence of discriminant and convergent validity, with validity diagonals averaging .33 (Inventory × Peer Ratings) and .39 (Self Ratings × Peer Ratings) against heterotrait-heteromethod control values averaging .14 and .16. While not intrinsically impressive, this picture is nonetheless better than most of the validity matrices here assembled. Note that the Self Ratings show slightly higher validity diagonal elevations than do the Inventory scores, in spite of the much greater length and undoubtedly higher reliability of the latter. In addition, a method factor seems almost totally lacking for the Self Ratings, while strongly present for the Inventory, so that the Self Ratings come off much the best if true trait variance is expressed as a proportion of total reliable variance (as Vernon [1958] suggests). The method factor in the STDCR Inventory is undoubtedly enhanced by scoring the same item in several scales, thus contributing correlated error variance, which could be reduced without loss of reliability by the simple expedient of adding more equivalent items and scoring each item in only one scale. It should be noted that Carroll makes explicit use of the comparison of the validity diagonal with the heterotrait-heteromethod values as a validity indicator.

Ratings in the Assessment Study of Clinical Psychologists

The illustrations of multitrait-multimethod matrices presented so far give a rather sorry picture of the validity of the measures of individual differences involved. The typical case shows an excessive amount of

TABLE 11

GUILFORD-MARTIN FACTORS STDCR AND RELATED RATINGS

(N=110)

	Inventory					Self Ratings					Peer Ratings				
	S	T	D	C	-R	S	T	D	C	-R	S	T	D	C	-R
Inventory															
S	(.92)														
T	.27	(.89)													
D	.62	.57	(.91)												
C	.36	.47	.90	(.91)											
-R	.69	.32	.28	−.06	(.89)										
Self Ratings															
S	*.57*	.11	.19	−.01	.53	( )									
T	.28	*.65*	.42	.26	.37	.26	( )								
D	.44	.25	*.53*	.45	.29	.31	.32	( )							
C	.31	.20	.54	*.52*	.13	.11	.21	.47	( )						
-R	.15	.30	.12	.04	*.34*	.10	.12	.04	.06	( )					
Peer Ratings															
S	*.37*	.08	.10	−.01	.38	*.42*	.02	.08	.08	.31	(.81)				
T	.23	*.32*	.15	.04	.40	.20	*.39*	.40	.21	.31	.37	(.66)			
D	.31	.11	*.27*	.24	.25	.17	.09	*.29*	.27	.30	.49	.38	(.73)		
C	.08	.15	.20	*.26*	−.05	.01	.06	.14	*.30*	.07	.19	.16	.40	(.75)	
-R	.21	.20	−.03	−.16	*.45*	.28	.17	.08	.01	*.56*	.55	.56	.34	−.07	(.76)

method variance, which usually exceeds the amount of trait variance. This picture is certainly not as a result of a deliberate effort to select shockingly bad examples: these are ones we have encountered without attempting an exhaustive coverage of the literature. The several unpublished studies of which we are aware show the same picture. If they seem more disappointing than the general run of validity data reported in the journals, this impression may very well be because the portrait of validity provided by isolated values plucked from the validity diagonal is deceptive, and uninterpretable in isolation from the total matrix. Yet it is clear that few of the classic examples of successful measurement of individual differences are involved, and that in many of the instances, the quality of the data might have been such as to magnify apparatus factors, etc. A more nearly ideal set of personality data upon which to illustrate the method was therefore sought in the multiple application of a set of rating scales in the assessment study of clinical psychologists (Kelly & Fiske, 1951).

In that study, "Rating Scale A" contained 22 traits referring to "behavior which can be directly observed on the surface." In using this scale the raters were instructed to "disregard any inferences about underlying dynamics or causes" (p. 207). The *S*s, first-year clinical psychology students, rated themselves and also their three teammates with whom they had participated in the various assessment procedures and with whom they had lived for six days. The median of the three teammates' ratings was used for the Teammate score. The *S*s were also rated on these 22 traits by the assessment staff. Our analysis uses the Final Pooled ratings, which were agreed upon by three staff members after discussion and review of the enormous amount of data and the many other ratings on each *S*. Unfortunately for our purposes, the staff members saw the ratings by Self and Teammates before making theirs, although presumably they were little influenced by these data because they had so much other evidence available to them. (See Kelly & Fiske, 1951, especially p. 64.) The Self and Teammate ratings represent entirely separate "methods" and can be given the major emphasis in evaluating the data to be presented.

In a previous analysis of these data (Fiske, 1949), each of the three heterotrait-monomethod triangles was computed and factored. To provide a multitrait-multimethod matrix, the 1452 heteromethod correlations have been computed especially for this report.[2] The full 66×66 matrix with its 2145 coefficients is obviously too large for presentation here, but will be used in analyses that follow. To provide an illustrative sample, Table 12 presents the interrelationships among five variables, selecting the one best representing each of the five recurrent factors discovered in Fiske's (1949) previous analysis of the monomethod matrices. (These were chosen without regard to their validity as indicated in the heteromethod blocks. Assertive—No. 3 reflected—was selected to represent Recurrent Factor 5 because Talkative had also a high

[2] We are indebted to E. Lowell Kelly for furnishing the V.A. assessment date to us, and to Hugh Lane for producing the matrix of intercorrelations.

In the original report the correlations were based upon 128 men. The present analyses were based on only 124 of these cases because of clerical errors. This reduction in N leads to some very minor discrepancies between these values and those previously reported.

TABLE 12

Ratings from Assessment Study of Clinical Psychologists
(N=124)

		Staff Ratings					Teammate Ratings					Self Ratings				
		A_1	B_1	C_1	D_1	E_1	A_2	B_2	C_2	D_2	E_2	A_3	B_3	C_3	D_3	E_3
Staff Ratings																
Assertive	A_1	(.89)														
Cheerful	B_1	.37	(.85)													
Serious	C_1	−.24	−.14	(.81)												
Unshakable Poise	D_1	.25	.46	.08	(.84)											
Broad Interests	E_1	.35	.19	.09	.31	(.92)										
Teammate Ratings																
Assertive	A_2	*.71*	.35	−.18	.26	.41	(.82)									
Cheerful	B_2	.39	*.53*	−.15	.38	.29	.37	(.76)								
Serious	C_2	−.27	−.31	*.43*	−.06	.03	−.15	−.19	(.70)							
Unshakable Poise	D_2	.03	−.05	.03	*.20*	.07	.11	.23	.19	(.74)						
Broad Interests	E_2	.19	.05	.04	.29	*.47*	.33	.22	.19	.29	(.76)					
Self Ratings																
Assertive	A_3	*.48*	.31	−.22	.19	.12	*.46*	.36	−.15	.12	.23	( )				
Cheerful	B_3	.17	*.42*	−.10	.10	−.03	.09	*.24*	−.25	−.11	−.03	.23	( )			
Serious	C_3	−.04	−.13	*.22*	−.13	−.05	−.04	−.11	*.31*	.06	.06	−.05	−.12	( )		
Unshakable Poise	D_3	.13	.27	−.03	*.22*	−.04	.10	.15	.00	*.14*	−.03	.16	.26	.11	( )	
Broad Interests	E_3	.37	.15	−.22	.09	*.26*	.27	.12	−.07	.05	*.35*	.21	.15	.17	.31	( )

loading on the first recurrent factor.)

The picture presented in Table 12 is, we believe, typical of the best validity in personality trait ratings that psychology has to offer at the present time. It is comforting to note that the picture is better than most of those previously examined. Note that the validities for Assertive exceed heterotrait values of both the monomethod and heteromethod triangles. Cheerful, Broad Interests, and Serious have validities exceeding the heterotrait-heteromethod values with two exceptions. Only for Unshakable Poise does the evidence of validity seem trivial. The elevation of the reliabilities above the heterotrait-monomethod triangles is further evidence for discriminant validity.

A comparison of Table 12 with the full matrix shows that the procedure of having but one variable to represent each factor has enhanced the appearance of validity, although not necessarily in a misleading fashion. Where several variables are all highly loaded on the same factor, their "true" level of intercorrelation is high. Under these conditions, sampling errors can depress validity diagonal values and enhance others to produce occasional exceptions to the validity picture, both in the heterotrait-monomethod matrix and in the heteromethod-heterotrait triangles. In this instance, with an N of 124, the sampling error is appreciable, and may thus be expected to exaggerate the degree of invalidity.

Within the monomethod sections, errors of measurement will be correlated, raising the general level of values found, while within the heteromethods block, measurement errors are independent, and tend to lower the values both along the validity diagonal and in the heterotrait triangles. These effects, which may also be stated in terms of method factors or shared confounded irrelevancies, operate strongly in these data, as probably in all data involving ratings. In such cases, where several variables represent each factor, none of the variables consistently meets the criterion that validity values exceed the corresponding values in the monomethod triangles, when the full matrix is examined.

To summarize the validation picture with respect to comparisons of validity values with other heteromethod values in each block, Table 13 has been prepared. For each trait and for each of the three heteromethod blocks, it presents the value of the validity diagonal, the highest heterotrait value involving that trait, and the number out of the 42 such heterotrait values which exceed the validity diagonal in magnitude. (The number 42 comes from the grouping of the 21 other column values and the 21 other row values for the column and row intersecting at the given diagonal value.)

On the requirement that the validity diagonal exceed all others in its heteromethod block, none of the traits has a completely perfect record, although some come close. Assertive has only one trivial exception in the Teammate-Self block. Talkative has almost as good a record, as does Imaginative. Serious has but two inconsequential exceptions and Interest in Women three. These traits stand out as highly valid in both self-description and reputation. Note that the actual validity coefficients of these four traits range from but .22 to .82, or, if we concentrate on the Teammate-Self block as most certainly representing independent methods, from but .31 to .46. While these are the best traits, it seems that most of the traits have far above

TABLE 13

VALIDITIES OF TRAITS IN THE ASSESSMENT STUDY OF CLINICAL PSYCHOLOGISTS, AS JUDGED BY THE HETEROMETHOD COMPARISONS

	Staff-Teammate			*Staff-Self*			*Teammate-Self*		
	Val.	Highest Het.	No. Higher	Val.	Highest Het.	No. Higher	Val.	Highest Het.	No. Higher
1. Obstructiveness*	.30	.34	2	.16	.27	9	.19	.24	1
2. Unpredictable	.34	.26	0	.18	.24	3	.05	.19	29
3. Assertive*	.71	.65	0	.48	.45	0	.46	.48	1
4. Cheerful*	.53	.60	2	.42	.40	0	.24	.38	5
5. Serious*	.43	.35	0	.22	.27	2	.31	.24	0
6. Cool, Aloof	.49	.48	0	.20	.46	10	.02	.34	36
7. Unshakable Poise	.20	.40	16	.22	.27	4	.14	.19	10
8. Broad Interests*	.47	.46	0	.26	.37	6	.35	.32	0
9. Trustful	.26	.34	5	.08	.25	19	.11	.17	9
10. Self-centered	.30	.34	2	.17	.27	6	−.07	.19	36
11. Talkative*	.82	.65	0	.47	.45	0	.43	.48	1
12. Adventurous	.45	.60	6	.28	.30	2	.16	.36	14
13. Socially Awkward	.45	.37	0	.06	.21	28	.04	.16	30
14. Adaptable*	.44	.40	0	.18	.23	10	.17	.29	8
15. Self-sufficient*	.32	.33	1	.13	.18	5	.18	.15	0
16. Worrying, Anxious*	.41	.37	0	.23	.33	5	.15	.16	1
17. Conscientious	.26	.33	4	.11	.32	19	.21	.23	2
18. Imaginative*	.43	.46	1	.32	.31	0	.36	.32	0
19. Interest in Women*	.42	.43	2	.55	.38	0	.37	.40	1
20. Secretive, Reserved*	.40	.58	5	.38	.40	2	.32	.35	3
21. Independent Minded	.39	.42	2	.08	.25	19	.21	.30	3
22. Emotional Expression*	.62	.63	1	.31	.46	5	.19	.34	10

Note.—Val. = value in validity diagonal; Highest Het. = highest heterotrait value; No. Higher = number of heterotrait values exceeding the validity diagonal.
* Trait names which have validities in all three heteromethod blocks significantly greater than the heterotrait-heteromethod values at the .001 level.

chance validity. All those having 10 or fewer exceptions have a degree of validity significant at the .001 level as crudely estimated by a one-tailed sign test.[3] All but one of the variables meet this level for the Staff-Teammate block, all but four for the Staff-Self block, all but five for the most independent block, Teammate-Self. The exceptions to significant validity are not parallel from column to column, however, and only 13 of 22 variables have .001 significant validity in all three blocks. These are indicated by an asterisk in Table 13.

This highly significant general level of validity must not obscure the meaningful problem created by the occasional exceptions, even for the best variables. The excellent traits of Assertive and Talkative provide a case in point. In terms of Fiske's original analysis, both have high loadings on the recurrent factor "Confident self-expression" (repre-

[3] If we take the validity value as fixed (ignoring its sampling fluctuations), then we can determine whether the number of values larger than it in its row and column is less than expected on the null hypothesis that half the values would be above it. This procedure requires the assumption that the position (above or below the validity value) of any one of these comparison values is independent of the position of each of the others, a dubious assumption when common methods and trait variance are present.

sented by Assertive in Table 12). Talkative also had high loadings on the recurrent factor of Social Adaptability (represented by Cheerful in Table 12). We would expect, therefore, both high correlation between them and significant discrimination as well. And even at the common sense level, most psychologists would expect fellow psychologists to discriminate validly between assertiveness (nonsubmissiveness) and talkativeness. Yet in the Teammate-Self block, Assertive rated by self correlates .48 with Talkative by teammates, higher than either of their validities in this block, .43 and .46.

In terms of the average values of the validities and the frequency of exceptions, there is a distinct trend for the Staff-Teammate block to show the greatest agreement. This can be attributed to several factors. Both represent ratings from the external point of view. Both are averaged over three judges, minimizing individual biases and undoubtedly increasing reliabilities. Moreover, the Teammate ratings were available to the Staff in making their ratings. Another effect contributing to the less adequate convergence and discrimination of Self ratings was a response set toward the favorable pole which greatly reduced the range of these measures (Fiske, 1949, p. 342). Inspection of the details of the instances of invalidity summarized in Table 13 shows that in most instances the effect is attributable to the high specificity and low communality for the self-rating trait. In these instances, the column and row intersecting at the low validity diagonal are asymmetrical as far as general level of correlation is concerned, a fact covered over by the condensation provided in Table 13.

The personality psychologist is initially predisposed to reinterpret self-ratings, to treat them as symptoms rather than to interpret them literally. Thus, we were alert to instances in which the self ratings were not literally interpretable, yet nonetheless had a diagnostic significance when properly "translated." By and large, the instances of invalidity of self-descriptions found in this assessment study are not of this type, but rather are to be explained in terms of an absence of communality for one of the variables involved. In general, where these self descriptions are interpretable at all, they are as literally interpretable as are teammate descriptions. Such a finding may, of course, reflect a substantial degree of insight on the part of these *S*s.

The general success in discriminant validation coupled with the parallel factor patterns found in Fiske's earlier analysis of the three intramethod matrices seemed to justify an inspection of the factor pattern validity in this instance. One possible procedure would be to do a single analysis of the whole 66×66 matrix. Other approaches focused upon separate factoring of heteromethods blocks, matrix by matrix, could also be suggested. Not only would such methods be extremely tedious, but in addition they would leave undetermined the precise comparison of factor-pattern similarity. Correlating factor loadings over the population of variables was employed for this purpose by Fiske (1949) but while this provided for the identification of recurrent factors, no single over-all index of factor pattern similarity was generated. Since our immediate interest was in confirming a pattern of interrelationships, rather than in describing it, an efficient short cut was available: namely to test the similarity of the sets of heter-

otrait values by correlation coefficients in which each entry represented the size values of the given heterotrait coefficients in two different matrices. For the full matrix, such correlations would be based upon the N of the $22 \times 21/2$ or 231 specific heterotrait combinations. Correlations were computed between the Teammate and Self monomethods matrices, selected as maximally independent. (The values to follow were computed from the original correlation matrix and are somewhat higher than that which would be obtained from a reflected matrix.) The similarity between the two monomethods matrices was .84, corroborating the factor-pattern similarity between these matrices described more fully by Fiske in his parallel factor analyses of them. To carry this mode of analysis into the heteromethod block, this block was treated as though divided into two by the validity diagonal, the above diagonal values and the below diagonal representing the maximally independent validation of the heterotrait correlation pattern. These two correlated .63, a value which, while lower, shows an impressive degree of confirmation. There remains the question as to whether this pattern upon which the two heteromethod-heterotrait triangles agree is the same one found in common between the two monomethod triangles. The intra-Teammate matrix correlated with the two heteromethod triangles .71 and .71. The intra-Self matrix correlated with the two .57 and .63. In general, then, there is evidence for validity of the intertrait relationship pattern.

Discussion

Relation to construct validity. While the validational criteria presented are explicit or implicit in the discussions of construct validity (Cronbach & Meehl, 1955; APA, 1954), this paper is primarily concerned with the adequacy of tests as measures of a construct rather than with the adequacy of a construct as determined by the confirmation of theoretically predicted associations with measures of other constructs. We believe that before one can test the relationships between a specific trait and other traits, one must have some confidence in one's measures of that trait. Such confidence can be supported by evidence of convergent and discriminant validation. Stated in different words, any conceptual formulation of trait will usually include implicitly the proposition that this trait is a response tendency which can be observed under more than one experimental condition and that this trait can be meaningfully differentiated from other traits. The testing of these two propositions must be prior to the testing of other propositions to prevent the acceptance of erroneous conclusions. For example, a conceptual framework might postulate a large correlation between Traits A and B and no correlation between Traits A and C. If the experimenter then measures A and B by one method (e.g., questionnaire) and C by another method (such as the measurement of overt behavior in a situation test), his findings may be consistent with his hypotheses solely as a function of method variance common to his measures of A and B but not to C.

The requirements of this paper are intended to be as appropriate to the relatively atheoretical efforts typical of the tests and measurements field as to more theoretical efforts. This emphasis on validational criteria appropriate to our present atheoretical level of test construction is not at all

incompatible with a recognition of the desirability of increasing the extent to which all aspects of a test and the testing situation are determined by explicit theoretical considerations, as Jessor and Hammond have advocated (Jessor & Hammond, 1957).

Relation to operationalism. Underwood (1957, p. 54) in his effective presentation of the operationalist point of view shows a realistic awareness of the amorphous type of theory with which most psychologists work. He contrasts a psychologist's "literary" conception with the latter's operational definition as represented by his test or other measuring instrument. He recognizes the importance of the literary definition in communicating and generating science. He cautions that the operational definition "may not at all measure the process he wishes to measure; it may measure something quite different" (1957, p. 55). He does not, however, indicate how one would know when one was thus mistaken.

The requirements of the present paper may be seen as an extension of the kind of operationalism Underwood has expressed. The test constructor is asked to generate from his literary conception or private construct not one operational embodiment, but two or more, each as different in research vehicle as possible. Furthermore, he is asked to make explicit the distinction between his new variable and other variables, distinctions which are almost certainly implied in his literary definition. In his very first validational efforts, before he ever rushes into print, he is asked to apply the several methods and several traits jointly. His literary definition, his conception, is now best represented in what his independent measures of the trait hold *distinctively* in common. The multitrait-multimethod matrix is, we believe, an important practical first step in avoiding "the danger . . . that the investigator will fall into the trap of thinking that because he went from an artistic or literary conception . . . to the construction of items for a scale to measure it, he has validated his artistic conception" (Underwood, 1957, p. 55). In contrast with the *single operationalism* now dominant in psychology, we are advocating a ***multiple operationalism***, a ***convergent operationalism*** (Garner, 1954; Garner, Hake, & Eriksen, 1956), a ***methodological triangulation*** (Campbell: 1953, 1956), an ***operational delineation*** (Campbell, 1954), a ***convergent validation***.

Underwood's presentation and that of this paper as a whole imply moving from concept to operation, a sequence that is frequent in science, and perhaps typical. The same point can be made, however, in inspecting a transition from operation to construct. For any body of data taken from a single operation, there is a subinfinity of interpretations possible; a subinfinity of concepts, or combinations of concepts, that it could represent. Any single operation, as representative of concepts, is equivocal. In an analogous fashion, when we view the Ames distorted room from a fixed point and through a single eye, the data of the retinal pattern are equivocal, in that a subinfinity of hexahedrons could generate the same pattern. The addition of a second viewpoint, as through binocular parallax, greatly reduces this equivocality, greatly limits the constructs that could jointly account for both sets of data. In Garner's (1954) study, the fractionation measures from a single method were equivocal—they could have been a function of the stimulus distance being fractionated, or they

could have been a function of the comparison stimuli used in the judgment process. A multiple, convergent operationalism reduced this equivocality, showing the latter conceptualization to be the appropriate one, and revealing a preponderance of methods variance. Similarly for learning studies: in identifying constructs with the response data from animals in a specific operational setup there is equivocality which can operationally be reduced by introducing transposition tests, different operations so designed as to put to comparison the rival conceptualizations (Campbell, 1954).

Garner's convergent operationalism and our insistence on more than one method for measuring each concept depart from Bridgman's early position that "if we have more than one set of operations, we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations" (Bridgman, 1927, p. 10). At the current stage of psychological progress, the crucial requirement is the demonstration of some convergence, not complete congruence, between two distinct sets of operations. With only one method, one has no way of distinguishing trait variance from unwanted method variance. When psychological measurement and conceptualization become better developed, it may well be appropriate to differentiate conceptually between Trait-Method Unit A_1 and Trait-Method Unit A_2, in which Trait A is measured by different methods. More likely, what we have called method variance will be specified theoretically in terms of a set of constructs. (This has in effect been illustrated in the discussion above in which it was noted that the response set variance might be viewed as trait variance, and in the rearrangement of the social intelligence matrices of Tables 4 and 5.) It will then be recognized that measurement procedures usually involve several theoretical constructs in joint application. Using obtained measurements to estimate values for a single construct under this condition still requires comparison of complex measures varying in their trait composition, in something like a multitrait-multimethod matrix. Mill's joint method of similarities and differences still epitomizes much about the effective experimental clarification of concepts.

The evaluation of a multitrait-multimethod matrix. The evaluation of the correlation matrix formed by intercorrelating several trait-method units must take into consideration the many factors which are known to affect the magnitude of correlations. A value in the validity diagonal must be assessed in the light of the reliabilities of the two measures involved: e.g., a low reliability for Test A_2 might exaggerate the apparent method variance in Test A_1. Again, the whole approach assumes adequate sampling of individuals: the curtailment of the sample with respect to one or more traits will depress the reliability coefficients and intercorrelations involving these traits. While restrictions of range over all traits produces serious difficulties in the interpretation of a multitrait-multimethod matrix and should be avoided whenever possible, the presence of different degrees of restriction on different traits is the more serious hazard to meaningful interpretation.

Various statistical treatments for multitrait-multimethod matrices might be developed. We have considered rough tests for the elevation

of a value in the validity diagonal above the comparison values in its row and column. Correlations between the columns for variables measuring the same trait, variance analyses, and factor analyses have been proposed to us. However, the development of such statistical methods is beyond the scope of this paper. We believe that such summary statistics are neither necessary nor appropriate at this time. Psychologists today should be concerned not with evaluating tests as if the tests were fixed and definitive, but rather with developing better tests. We believe that a careful examination of a multitrait-multimethod matrix will indicate to the experimenter what his next steps should be: it will indicate which methods should be discarded or replaced, which concepts need sharper delineation, and which concepts are poorly measured because of excessive or confounding method variance. Validity judgments based on such a matrix must take into account the stage of development of the constructs, the postulated relationships among them, the level of technical refinement of the methods, the relative independence of the methods, and any pertinent characteristics of the sample of *S*s. We are proposing that the validational process be viewed as an aspect of an ongoing program for improving measuring procedures and that the "validity coefficients" obtained at any one stage in the process be interpreted in terms of gains over preceding stages and as indicators of where further effort is needed.

The design of a multitrait-multimethod matrix. The several methods and traits included in a validational matrix should be selected with care. The several methods used to measure each trait should be appropriate to the trait as conceptualized. Although this view will reduce the range of suitable methods, it will rarely restrict the measurement to one operational procedure.

Wherever possible, the several methods in one matrix should be completely independent of each other: there should be no prior reason for believing that they share method variance. This requirement is necessary to permit the values in the heteromethod-heterotrait triangles to approach zero. If the nature of the traits rules out such independence of methods, efforts should be made to obtain as much diversity as possible in terms of data-sources and classification processes. Thus, the classes of stimuli *or* the background situations, the experimental contexts, should be different. Again, the persons providing the observations should have different roles *or* the procedures for scoring should be varied.

Plans for a validational matrix should take into account the difference between the interpretations regarding convergence and discrimination. It is sufficient to demonstrate convergence between two clearly distinct methods which show little overlap in the heterotrait-heteromethod triangles. While agreement between several methods is desirable, convergence between two is a satisfactory minimal requirement. Discriminative validation is not so easily achieved. Just as it is impossible to prove the null hypothesis, or that some object does not exist, so one can never establish that a trait, as measured, is differentiated from all other traits. One can only show that this measure of Trait A has little overlap with those measures of B and C, and no dependable generalization beyond B and C can be made. For example, social poise could probably

be readily discriminated from aesthetic interests, but it should also be differentiated from leadership.

Insofar as the traits are related and are expected to correlate with each other, the monomethod correlations will be substantial and heteromethod correlations between traits will also be positive. For ease of interpretation, it may be best to include in the matrix at least two traits, and preferably two sets of traits, which are postulated to be independent of each other.

In closing, a word of caution is needed. Many multitrait-multimethod matrices will show no convergent validation: no relationship may be found between two methods of measuring a trait. In this common situation, the experimenter should examine the evidence in favor of several alternative propositions: (*a*) Neither method is adequate for measuring the trait; (*b*) One of the two methods does not really measure the trait. (When the evidence indicates that a method does not measure the postulated trait, it may prove to measure some other trait. High correlations in the heterotrait-heteromethod triangles may provide hints to such possibilities.) (*c*) The trait is not a functional unity, the response tendencies involved being specific to the nontrait attributes of each test. The failure to demonstrate convergence may lead to conceptual developments rather than to the abandonment of a test.

SUMMARY

This paper advocates a validational process utilizing a matrix of intercorrelations among tests representing at least two traits, each measured by at least two methods. Measures of the same trait should correlate higher with each other than they do with measures of different traits involving separate methods. Ideally, these validity values should also be higher than the correlations among different traits measured by the same method.

Illustrations from the literature show that these desirable conditions, as a set, are rarely met. Method or apparatus factors make very large contributions to psychological measurements.

The notions of convergence between independent measures of the same trait and discrimination between measures of different traits are compared with previously published formulations, such as construct validity and convergent operationalism. Problems in the application of this validational process are considered.

REFERENCES

AMERICAN PSYCHOLOGICAL ASSOCIATION. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull., Suppl.*, 1954, **51**, Part 2, 1–38.

ANDERSON, E. E. Interrelationship of drives in the male albino rat. I. Intercorrelations of measures of drives. *J. comp. Psychol.*, 1937, **24**, 73–118.

AYER, A. J. *The problem of knowledge.* New York: St Martin's Press, 1956.

BORGATTA, E. F. Analysis of social interaction and sociometric perception. *Sociometry*, 1954, **17**, 7–32.

BORGATTA, E. F. Analysis of social interaction: Actual, role-playing, and projective. *J. abnorm. soc. Psychol.*, 1955, **51**, 394–405.

BRIDGMAN, P. W. *The logic of modern physics.* New York: Macmillan, 1927.

BURWEN, L. S., & CAMPBELL, D. T. The generality of attitudes toward authority and nonauthority figures. *J. abnorm. soc. Psychol.*, 1957, **54**, 24–31.

CAMPBELL, D. T. *A study of leadership among submarine officers.* Columbus: Ohio State Univer. Res. Found., 1953.

CAMPBELL, D. T. Operational delineation of "what is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167–174.

CAMPBELL, D. T. *Leadership and its effects upon the group.* Monogr. No. 83. Columbus: Ohio State Univer. Bur. Business Res., 1956.

CARROLL, J. B. Ratings on traits measured by a factored personality inventory. *J. abnorm. soc. Psychol.*, 1952, **47**, 626–632.

CHI, P.-L. Statistical analysis of personality rating. *J. exp. Educ.*, 1937, **5**, 229–245.

CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.

CRONBACH, L. J. *Essentials of psychological testing.* New York: Harper, 1949.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

FEIGL, H. The mental and the physical. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science.* Vol. II. *Concepts, theories and the mind-body problem.* Minneapolis: Univer. Minnesota Press, 1958.

FISKE, D. W. Consistency of the factorial structures of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, **44**, 329–344.

GARNER, W. R. Context effects and the validity of loudness scales. *J. exp. Psychol.*, 1954, **48**, 218–224.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.

JESSOR, R., & HAMMOND, K. R. Construct validity and the Taylor Anxiety Scale. *Psychol. Bull.*, 1957, **54**, 161–170.

KELLEY, T. L., & KREY, A. C. *Tests and measurements in the social sciences.* New York: Scribner, 1934.

KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. Michigan Press, 1951.

LOEVINGER, J., GLESER, G. C., & DUBOIS, P. H., Maximizing the discriminating power of a multiple-score test. *Psychometrika*, 1953, **18**, 309–317.

LORGE, I. Gen-like: Halo or reality? *Psychol. Bull.*, 1937, **34**, 545–546.

MAYO, G. D. Peer ratings and halo. *Educ. psychol. Measmt*, 1956, **16**, 317–323.

STRANG, R. Relation of social intelligence to certain other factors. *Sch. & Soc.*, 1930, **32**, 268–272.

SYMONDS, P. M. *Diagnosing personality and conduct.* New York: Appleton-Century, 1931.

THORNDIKE, E. L. A constant error in psychological ratings. *J. appl. Psychol.*, 1920, **4**, 25–29.

THORNDIKE, R. L. Factor analysis of social and abstract intelligence. *J. educ. Psychol.*, 1936, **27**, 231–233.

THURSTONE, L. L. *The reliability and validity of tests.* Ann Arbor: Edwards, 1937.

TRYON, R. C. Individual differences. In F. A. Moss (Ed.), *Comparative Psychology.* (2nd ed.) New York: Prentice-Hall, 1942. Pp. 330–365.

UNDERWOOD, B. J. *Psychological research.* New York: Appleton-Century-Crofts, 1957.

VERNON, P. E. Educational ability and psychological factors. Address given to the Joint Education-Psychology Colloquium, Univer. of Illinois, March 29, 1957.

VERNON, P. E. *Educational testing and test-form factors.* Princeton: Educational Testing Service, 1958. (Res. Bull. RB-58-3.)

Received June 19, 1958.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 2, 1959

THE IMPORTANCE OF TIME AS STIMULUS OF CONDITIONED REFLEX ACTIVITY[1]

A. S. DMITRIEV AND A. M. KOCHIGINA
Iaroslavl'

Translated from *Uspekhi Sovremennoi Biologii*

Investigations extending over many years on the conditioned reflex activity of the higher animals and man have shown that "innumerable variations in both the external and internal media of the organism, each of which is reflected in certain states of the cortical nerve cells, may themselves become individual conditioned stimuli" (Pavlov, 1947, p. 51).

It is quite natural that time, as one of the fundamental attributes of the existence of matter, and being closely linked with all variations in the external and internal media of the organism, should be very frequently found to be one of the components of the conditioned stimulus, emerging in the role of an exciter of conditioned reflex activity *sui generis*. The conditioned stimulus is not, of course, time merely as time, but a change predestined to occur at a definite time in the state of the cortical cells, a state induced by the action of the external or internal medium of the organism. In consequence of the linking of these changes with time, the conditions are created for the reflection of time as an objective reality in the cerebral cortex and for the formation of conditioned reflexes to certain segments of time, as was demonstrated experimentally by the investigations of I. P. Pavlov and his coworkers.

The formation of conditioned reflexes to time plays an important part in the systemic activity of the cerebral cortex, in the development of definite periodicity in physiological functions, and in the establishment of rhythmical pattern reactions in the working activity of man. By virtue thereof the question of the cortical mechanisms involved in the process of formation of conditioned reflexes to time acquires considerable theoretical and practical importance. The present work is concerned with the analysis and generalization of findings given in the literature and the results of our own investigations.

I

The fact that conditioned reflexes are formed to time was first established in Pavlov's laboratory, in the experiments of G. P. Zelenyi (1907), and later, those of K. N. Krzhishkovskii (1908), on the establishment of conditioned salivary reflexes in dogs, when the conditioned stimuli were administered at equal intervals of time.

G. P. Zelenyi, working on the formation of conditioned salivary reflexes to various sound stimuli, administered a combination of sound stimulus and feeding every 10 minutes. After a number of repetitions, it was noted that conditioned reflex salivation occurred regularly at the end of the 10-minute interval, even in the absence of conditioned stimulation.

[1] This article was first published in *Uspekhi sovremennoi biologii*, 1955, 40, 31–51. It was translated by the Pergamon Institute for the Russian Scientific Translation Program of the National Institutes of Health, U. S. Public Health Service. Reprints of this article may be obtained by writing directly to the Editor of the *Psychological Bulletin*.

In the work of K. N. Krzhishkovskii on an investigation of conditioned inhibition it was established that the application of inhibition constantly at the nineteenth to twentieth minute led to the following events: the conditioned stimulus always proved inactive at the nineteenth to twentieth minute, but manifested its effect at the thirty-third to thirty-fourth minute. Krzhishkovskii observed this phenomenon over a period of six months. Wishing to determine whether time was the factor determining inhibition under these circumstances, he set up experiments for the establishment of a conditioned salivary reflex when a combination of the unconditioned stimulation (solution of hydrochloric acid) with the conditioned stimulus was administered constantly every 10–13 minutes. It was found that "if the animal's oral cavity was stimulated at equal intervals of time, then,"—in the author's words—"with the passage of time such a state developed that at the appropriate moment, even in the absence of any obvious stimulations, a flow of saliva began, and frequently also there was a characteristic motor reaction."

The importance of time as a factor in the establishment of conditioned reflexes has also been noted by other investigators in Pavlov's laboratory in works referable to the same period. Thus, observations by I. P. Pimenov (1907), I. V. Zavadskii (1908), F. S. Grossman (1909), and V. M. Dobrovol'skii (1911) showed that, in the case of trace and delayed reflexes, reaction to the conditioned stimulation supervened only after a certain time from the commencement of its action, that is, after the lapse of the segment of time which habitually elapsed before reinforcement was administered. In other words, the conditioned reflex was formed not simply to a particular stimulus, but to stimulus plus a certain segment of time.

These various investigations justified consideration of the question of time as a stimulus of signal significance, and of the power of the cortical cells to reckon time. And even at that time Pavlov, on the basis of existing factual material, advanced a hypothesis on the mechanism of time reckoning in the central nervous system.

"It may be thought," he wrote, "that intensity analysis, in part at least, is a basic element in the measurement of time by animals. One can speculate whether any external agent of uniform, constant strength acts on the particular analyser of the animal, and whether there is gradual fading in the nerve cells of the residuum, the trace of actual stimulation which has ceased; each intensity of the stimulated state of the cell, at each separate moment of time, is an individual element, differing from both all preceding and all following stages of intensity. Time is perhaps measured by these elements as units, and every moment of time recorded in the nervous system" (1947, pp. 115–116).

Investigations dealing specially with the role of time in the formation of conditioned reflexes were first carried out by Iu. P. Feokritova (1912) at Pavlov's suggestion. "We proposed to determine," she wrote, "whether time . . . could be linked, like all phenomena in the external environment, with the activity of the salivary gland, and whether it could serve as a specific salivary stimulus. If this proved to be the case, we wished to determine how rapidly the link between a definite time and the work of the salivary gland could be established, and thereafter to study to the fullest possible extent the

characteristic features of the new reflex to time" (1912, p. 17).

With these ends in view, Iu. P. Feokritova established conditioned salivary reflexes in dogs, repeating the combination of conditioned stimulus (sound of the metronome) and unconditioned (feeding of meat-biscuit powder or the introduction of hydrochloric acid into the mouth) at predetermined intervals of time (in the case of one dog, every 30 minutes; for another every 15 minutes; and in the case of a third, every 10 minutes). In a number of the experiments only one reinforcement was administered —at the same intervals of time.

A stable reflex to time was formed after 200–230 repetitions (the number varied for the different animals), and was evidenced by the fact that after a certain interval of time, and just before the time for the next reinforcement, saliva secretion began. The author also noted that, before the formation of the stable conditioned reflex to time, there was irregular salivation in the intervals between reinforcements. Later, salivation began to be limited in time to the second half of the interval, and ultimately approximated more and more closely to the moment of reinforcement. Salivation in the interval between reinforcements was, however, not infrequent (intersignal reaction) even after the formation of a stable conditioned reflex to time.

According to Feokritova's observations, the conditioned reflex to "pure time" was formed more rapidly than when time was combined with an auditory stimulus. She explains this on the grounds that time, as an independent stimulus, is very weak, and that the auditory stimulus overshadows its action.

Examining the effect of the metronome in combination with different intervals of time (except that to which the conditioned reflex was formed), Feokritova found that the effect of the metronome, when not associated with a definite interval of time, was negligible. Thus, in the dog in which a conditioned reflex to a 30-minute interval was established the operation of the metronome at the fourteenth minute induced secretion of only one drop of saliva, whereas at the thirtieth minute the action of the same metronome was accompanied by the secretion of 6–11 drops. The author concludes from this that "in our summated reflex, time is the more active excitant and consequently also the main component, and not the metronome" (1912, p. 57).

By her specially designed experiments Feokritova established that the differentiation of time can be brought to such a degree of accuracy that, for example, when a conditioned reflex has been established with a 30-minute interval, salivation is timed to occur exactly at the thirtieth minute, and is completely absent a minute earlier.

In her next series of experiments Feokritova tested the action of various extraneous stimuli (whistle, ventilator, gramophone) on the course of the conditioned reflex to time. Applied in the middle of the interval between reinforcements, such stimuli did not, as a rule, occasion any changes whatever in the course of the reflex. When they were used simultaneously with the metronome, then the conditioned reflex to time was inhibited (although not in all dogs, and not in the same degree). Applied a minute before the metronome, and ending 20–30 seconds before it began to act, they sometimes inhibited and sometimes intensified the conditioned reflex to time. As the reflex became stable, the effect of extraneous stim-

uli became attenuated, or even disappeared. When powerful extraneous stimuli were allowed to act and when any sudden change was introduced into the course or setting of the experiments, time differentiation was completely destroyed.

In this connection it is fitting to recall the observations of the author on the effect of a somnolent state, which the animals frequently developed in the course of the experiments, on the course of the conditioned reflex. Feokritova noted that a somnolent state not only failed to weaken the conditioned reflex to time, but even increased the differentiating power of the animal.

Feokritova's observations showed that extinction of the conditioned reflex to time occurred suddenly. Noting the similarity in this respect to the extinction of trace reflexes, the author assigns the conditioned reflex to time to the group of trace reflexes.

Feokritova concluded from her investigations that time can be an excitant of conditioned reflex activity in the salivary gland, and that the formation of a conditioned reflex to time was subject to the same laws as the formation of conditioned reflexes to other stimuli.

In offering an explanation of the mechanism for the reckoning of time by the nervous system, Feokritova wrote: "It is known that, after every stimulation, there remains in the cerebral cortex the trace of a whole series of states of stimulation of the nerve cell, gradually diminishing in intensity. At every individual moment, in each short interval of time, the intensity of the nerve cell stimulation will be different from what it is at any other moment. If we take the intensity of nerve cell stimulation during each short interval of time as an independent unit of stimulation, it can be stated that the exciter of the salivary centre in the case of a time reflex is that degree of intensity that always coincides in time with the act of eating" (1912, p. 162–163).

The work of M. M. Stukova (1914) was a direct continuation of Feokritova's investigations, and she confirmed the fact that a conditioned reflex could be established to time (a 20-minute interval in combination with mechanical skin stimulation), the time reflex being formed very rapidly in her experiments (after 65 repetitions). In dogs which had been transferred to her from Feokritova, Stukova noted the complete retention of the conditioned reflexes and time differentiation established earlier.

In her own experiments Stukova tested the effect of quite a range of different factors on the course of the conditioned reflex to time. Above all, she tested the effect of faradic currents of different strengths. Her observations demonstrated that the effect depended on the current strength and the characteristics of the animals' nervous systems. In the dog with well-developed inhibition processes currents of various strengths had no appreciable effect on the conditioned reflex, and only at the commencement of current action during the first application was some very transient disturbance of time differentiation observed. In excitable dogs inhibition of the conditioned reflex to time and disturbance of the correct reckoning of time were observed as a result of the action of faradic current, particularly if strong. A powerful current, acting momentarily, inhibited the reflex to time in all animals. Thus, like Feokritova, Stukova noted the inhibitory effect of change in the setting and course of the experiment on the conditioned reflex.

In her next series of experiments Stukova examined the effect of caffeine (0.05 gm.) and of cocaine (0.02–0.03 gm.) on conditioned reflexes to time. After subcutaneous injection of caffeine, all the animals showed (after 8–11 minutes) increased excitability, disturbance in the reckoning of time, and disinhibition of time differentiation. Under the action of cocaine, in addition to increased excitability, the appearance of continuous salivation in the intervals between reinforcements was observed. The reckoning of time by the nervous system was upset, and differentiation was disordered.

Stukova established in special experiments that the conditioned reflex to time appears (although varying in the degree of its expression) in response to any of the components in the summated stimulus (*e.g.*, to sight or smell of food only, without the reinforcement, and also to the action of the metronome alone or of the mechanical skin stimulation alone).

"The reckoning of time by the animals," Stukova concluded, "is possible, not only from a compound stimulation, that is, from the sum of many stimuli, as in our case (natural alimentary and artificial conditioned reflexes), but also from any one of these, taken separately" (p. 136).

The investigation of conditioned time reflexes was continued by V. S. Deriabin (1916) on the same animals. When he started work with these animals, he first of all noted that in all the dogs the conditioned reflexes to time produced earlier were restored at the first attempt and immediately attained their former magnitude.

Deriabin then tested the effect of a three-month interruption in the work. He demonstrated in this way that, with the lapse of three months, all the dogs manifested disorder of time reckoning.

Change in the experimental setting (particularly isolation of the animal from the experimenter) had the same effect on the reflex. In this case the time reflex and time differentiations disappeared, to reappear only after prolonged repetition.

Deriabin made a detailed investigation of the effects of sodium bromide and chloral hydrate on the reflex to time; these substances were injected into the animal's rectum, dissolved in distilled water. Sodium bromide, 2 gm., had no effect on the magnitude of the conditioned reflex to time: "The intensification of the processes of internal inhibition," wrote the author in this connection, "which is usually seen with sodium bromide in this quantity, did not lead to increase in the power of the dog's central nervous system to discriminate time. Time differentiation remained unchanged" (p. 94). Chloral hydrate, 0.5 gm. to 2.0 gm., had no appreciable effect on the conditioned reflex to time, but in doses of 4–5 gm. it had a suppressing action on the reflex, although the degree of suppression differed in different animals.

This author's findings relative to changes in the conditioned reflex to time after destruction of individual parts of the cerebral cortex are of considerable interest. In one dog in which a conditioned reflex to the action of a "kololka" (mechanical skin stimulator) in combination with a 20-minute interval of time had been established the cerebral region of the cutaneous analyser on one side was destroyed. In another dog in which a conditioned reflex to the sound of the metronome in combination with a time interval of 15 minutes had been produced, the cerebral region of the auditory analyser was removed on

both sides. In both cases the author noted temporary disappearance after the operation of the conditioned reflex to the corresponding stimulus ("kololka," metronome) in combination with time. With the lapse of a certain period, however, the conditioned reflex to time was restored completely: the conditioned reflex to "pure time" was quite clearly expressed in one dog seven days, and in the other two days after the operation.

In his own work Deriabin then examined the question of whether the nervous system of the dog is capable of discriminating between the time properties of sound and its other properties. For this purpose he set up experiments in which conditioned reflexes were established to various sounds (tuning-fork, gurgling sound, whistle) with rhythmical intervals (two seconds' sound—two seconds' interval). Differentiation of the basic rhythm from rhythms of the same sound, but with longer or shorter sounds and intervals, was then achieved. These experiments showed that rhythm differentiation of this type, established to one sound stimulus (whistling), appeared immediately when other sound stimuli (buzzer, tuning-fork) were administered.

The author therefore concluded that "the dog is capable of distinguishing the time properties of an auditory stimulus from other properties of sound, and is guided by them in its physiological activity," and that the nervous system of the dog "is capable of correlating similar time properties of several sound stimulations, as something common to the various types of sounds" (p. 141).

We see, therefore, that the works of Iu. P. Feokritova, M. M. Stukova, and V. S. Deriabin have provided a simple experimental basis for Pavlov's hypothesis on time as a true excitant of conditioned reflex activity in animals.

"It can be stated," wrote Pavlov in connection with Feokritova's experiments, "that in this particular case time was the conditioned stimulus" (1947, p. 50). This stimulus "is in no way less real," he pointed out, "than all the preceding stimuli" (1947, p. 49).

"Physiologically, how are we to understand time as a conditioned stimulus?" asked Pavlov. In answering this question, he started from the fact that in ordinary life we note time with the aid of certain cyclical phenomena, the setting and rising of the sun, the movements of hands on the dials of clocks, etc. And in our own bodies there are not a few of these cyclical phenomena. In the course of a day the brain receives stimulations, becomes exhausted and then is restored again. The digestive canal is periodically filled with food, and periodically emptied. And in that every state of an organ can be reflected in the cerebral hemispheres, we have there a basis for distinguishing one moment of time from another. Let us take short intervals of time. When stimulation has just been administered it is felt very acutely. When we enter a room in which there is some odor or other, we at first feel it very strongly, but afterwards less and less. Under the influence of stimulation the state of the nerve cell undergoes a series of changes. It is exactly the same in the converse case. When stimulation ceases, this is felt very acutely at first, and then less and less, until finally we are unaware of it. This again means that there are a number of different states of the nerve cell. Cases of reflexes to interruption of the stimulus

and trace reflexes, as well as cases of reflexes to time can be explained from the same point of view. In the experiment referred to (an experiment of Feokritova) the animal was fed periodically, and in association therewith a number of organs manifested a certain activity, that is, they underwent certain successive changes. All this was registered in the cerebral hemispheres, was received by them, and the definite moment of these changes became a conditioned stimulus" (1947, p. 50–51).

After the work of Feokritova, Stukova, and Deriabin, salivary conditioned reflexes to time were examined by a number of authors. These investigations, as well as giving greater definition to the conditions attending the formation of reflexes to time, demonstrated their importance in the systemic activity of the cerebral cortex.

Thus, F. D. Vasilenko (1932), working with a pattern of conditioned food stimuli, repeated at identical intervals of time, noted increase from experiment to experiment in the conditioned reflex reaction to one of the weak stimuli in the pattern (light). The author established that this increase in the secretory effect to light was the result of the summation of the conditioned reflex to light with the reflex to time which was being developed.

F. P. Maiorov (1933) also describes the formation of a conditioned reflex to time in relation to the action of a pattern. The author, having established a conditioned salivary reflex to the sound of the metronome (60 beats per minute) and a motor conditioned reflex to a different metronome frequency (120 beats per minute), began to alternate the action of these stimuli at equal intervals of time (six minutes). Comparatively soon (after nine repetitions) the author was able to establish a conditioned reflex to time, in which at first, in complete independence of the nature of the conditioned stimulation, both salivation and motor reaction appeared towards the end of the six-minute interval. Later, however, the conditioned reflex to time became differentiated; towards the end of the sixth minute, before the operation of the metronome (120 per minute), the motor reaction appeared, and towards the end of the sixth minute before the operation of the metronome (60 beats per minute), salivation began.

E. G. Vatsuro (1948), employing a pattern of stimuli, established in dogs, over a period of several months without interruption, a conditioned reflex to time, which manifested itself in the secretion of saliva 15–30 seconds before the beginning of the action of the conditioned stimulus, but subsequently the moment of the commencement of salivation coincided more and more closely with the moment of action of the conditioned stimulus. On other days feeding of the animal was carried out at the same intervals of time as in the preceding experiments, but without administration of the conditioned stimuli. It was found that in this case the reflex to time reproduced the picture of the conditioned pattern of reaction exactly. "The exactness of the reproduction," wrote the author, "reached the point of complete identity of the ratio of the values of the individual reflexes in the system with the corresponding reflexes to time" (1948, p. 13).

The importance of the time factor in the systemic activity of the cerebral cortex has been noted in their respective works by P. S. Kupalov (1929, 1931, 1933), E. A. Asratian (1938) and others.

The features in the formation of

conditioned reflexes to time, originally discovered in experiments on dogs by the salivary secretion method, have been given greater definition by investigations on the motor conditioned reflexes to time in these animals.

One of the first descriptions of the motor-defence conditioned reflex to time in dogs is found in the work of I. S. Beritov (1932). In experiments on dogs he repeated electrical stimulation, causing a defensive reflex flexion of a front limb, every five minutes and found that a conditioned reflex to time was formed after 40 repetitions. This reflex manifested itself in the fact that a minute before the next stimulation was due, the animal (which up to this point had been in a somnolent state) woke up, shook its head, and raised the limb.

In the literature we meet with very diverse findings on the question of the rate of, and conditions for, the formation of conditioned motor reflexes to time in dogs. Thus, in the work of L. S. Gambarian (1952) there is evidence of the very rapid formation of the motor-defence conditioned reflex to time in the case of a three-minute intersignal interval (unfortunately, the author, whose investigation had different aims, did not dwell in detail on his description of the process of formation of the conditioned reflex to time). On the other hand, O. P. Bolotina (1952a) states that with a three-minute interval the conditioned reflex to time is formed much more slowly, and sometimes fails completely to develop. Thus, having established a conditioned reflex to pressure with a forepaw on the pedal of a special apparatus in dogs, and repeating it with food reinforcement at 10-minute intervals, Bolotina noted the formation of a conditioned reflex to time only after 18 experiments (180 combinations). The attempt to transfer the conditioned reflex for the 10-minute intervals to shorter intervals (1–3 minutes) was successful only in rare cases. The author noted the formation of an unstable conditioned reflex to a two-minute interval after 52 experiments (520 combinations) in only one of three dogs. In the other dogs this was not achieved; a stable conditioned reflex could not be formed even to a three-minute interval, despite a large number of repetitions (344). In these experiments Bolotina noted only the appearance of a large number of intersignal reactions which, with repetition of the experiments, did not diminish in number, and did not become concentrated in the second half of the intersignal interval, as occurred in the course of the formation of the conditioned reflex to the 10-minute interval. The difficulty, and in a number of cases the impossibility, of forming conditioned reflexes to short intervals of time is due, in the opinion of the author, to the fact that in the case of a very short interval of time the traces of stimulation fail, against a background of alimentary excitation, to become concentrated, but, being summated, radiate through the cortex, and manifest themselves as a mass of intersignal pressures (1952a, p. 32).

According to the observations of Bolotina (1952a), a motor conditioned reflex to time is formed more readily with a compound stimulus than to "pure" time. Extinction of the conditioned reflex to time was reached after three–five applications without reinforcement. Examining the effect of extraneous stimuli on the conditioned reflex to time, the author found that the operation of extraneous stimuli immediately before the reinforcement or in the first seconds of its application caused inhibition of the conditioned reflex to

time in a number of cases, but their operation anywhere in the middle part of the intersignal interval usually failed to cause any appreciable changes in the conditioned reflex to time. Bolotina's observations indicate that changes in the experimental setting or interruptions in the work produced very marked disturbances in the conditioned reflex to time. Special experiments made by this author showed that hunger disturbed the animal's conditioned reflex to time, the manifestations being the appearance of intersignal motor reactions.

In another series of experiments Bolotina (1953) investigated the effect of bromide and caffeine on motor conditioned reflexes to time in dogs. This work showed that bromide in doses of 0.5–1.0 gm. hastened the formation of the conditioned reflex to time, and at the same time created conditions for the formation of a conditioned reflex to shorter intervals of time (two minutes), to which previously a conditioned reflex to time could not be established. When the conditioned time reflex was disturbed (as, for example, in attempts to transfer it to shorter intervals of time), the use of bromide in the same dose promoted its restoration and stabilization. In doses of 2–3 gm., bromide intensified the conditioned reflex to time. The use of bromide in combination with caffeine had a more pronounced positive effect on conditioned reflexes to time in dogs.

Motor-defence conditioned reflexes to time have been studied in detail by A. M. Kochigina in our laboratory. Her experiments showed that a motor-defence conditioned reflex to "pure" time (electrical stimulation repeated every five minutes) was formed very slowly, passing through certain stages in the course of its development. The process of conditioned time reflex formation began with the appearance of intersignal motor reactions, the number of which increased every time, reaching 50–60 in each five-minute interval. Within the limits of the five-minute interval the intersignal reactions were irregularly distributed, as the dotted line curve in Fig. 1 shows. This curve has a markedly undulant character, which means that the groups of intersignal reactions alternated with pauses of longer or shorter duration, in which the intersignal reactions were absent or scanty. This initial stage in the formation of the conditioned time reflex we have called the stage of generalized conditioned reflex to time.

In the next stage the number of intersignal reactions begins to diminish, and they become less and less frequent in the first half of the five-minute interval until they are no longer observed in this period, being concentrated in the second half of the interval, and increasing in number as the moment for the next reinforcement approaches. The change in the number of intersignal reactions occurring in the five-minute interval during this stage is shown by the continuous line curve in Fig. 1. We see that it is shorter by half than the dotted line, being displaced into the second half of the interval, and shows a notable rise towards the moment of the next reinforcement. We have termed this stage in the formation of the conditioned reflex to time the stage of formation of the differentiated conditioned reflex to time.

Subsequently, the number of intersignal reactions continues to diminish until they disappear completely. Only in the few seconds preceding the moment of the next reinforcement is there the occasional appearance of a

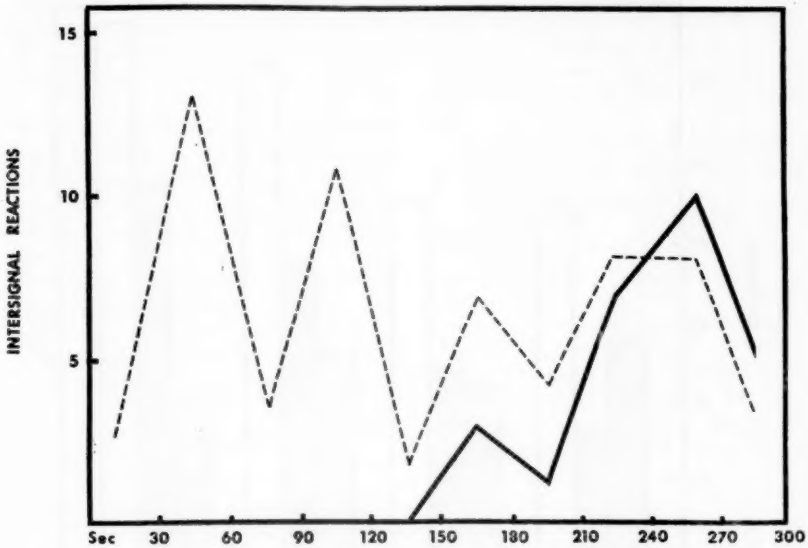

FIG. 1. DETAILS OF THE APPEARANCE OF INTERSIGNAL REACTIONS AT VARIOUS STAGES IN THE FORMATION OF A MOTOR-DEFENSE CONDITIONED REFLEX TO TIME (DOG, ROZKA).

Time is plotted on the abscissa from the moment of application of the conditioned stimulus, and on the ordinate, the number of intersignal reactions in every 30-second segment of the five-minute intersignal interval. The dotted line shows the distribution of intersignal reactions in one of the five-minute intervals in experiment No. 4 of 13 October 1953, and the continuous line, those in a five-minute interval from experiment No. 27 of 14 November 1953.

motor reaction. Thus, a more or less differentiated conditioned reflex to time has been formed. It was formed after 90–135 repetitions. The accuracy of the time reckoning after the formation of the differentiated conditioned reflex to time (expressed as the ratio of the time to the appearance of motor reactions to the total duration of the intersignal interval) ranged from 80 to 93 per cent in our dogs.

The duration of the individual stages and their relationship to one another in the process of the formation of a conditioned reflex to time varied in the different animals, as will be readily seen from Fig. 2. The differences in the duration of the second stage were particularly noticeable, and a certain parallelism was observed between the duration of this stage and the total time required for the formation of the differentiated conditioned reflex to time.

The motor-defence conditioned reflex to time was formed in the dogs more rapidly to the compound stimulus (sound plus the five-minute interval) than to "pure" time (after 64 and 97 repetitions respectively). During the process of the formation of the conditioned time reflex to the compound stimulus exactly the same stages were observed, but the number of intersignal reactions under these experimental conditions was considerably less. The accuracy of time

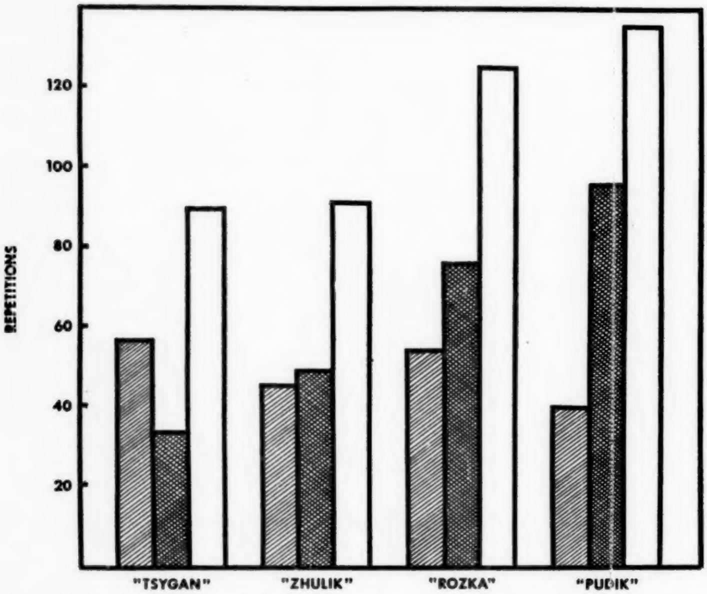

FIG. 2. DURATION OF THE STAGES IN THE PROCESS OF FORMATION OF A MOTOR-DEFENCE CONDITIONED REFLEX TO TIME IN DOGS.

The columns with lines in one diagonal show the duration (in number of repetitions) of the initial generalized stage of the conditioned reflex to time, the hatched columns that of the stage of the differentiated time reflex, and the white columns the total time required for the formation of the differentiated reflex to time.

reckoning after the formation of the differentiated conditioned reflex to the compound stimulus was the same as for "pure" time.

In the dogs the conditioned reflex to time was extinguished very rapidly (after 4–5 unreinforced applications). After a 10-day interruption, the conditioned time reflex was restored after 10–15 repetitions, the picture of restoration of the conditioned reflex recalling that of its formation, but with a more rapid sequence of stages.

We were able to confirm the findings of our predecessors relative to the effect of extraneous stimuli, bromide and caffeine, on the conditioned reflex to time. According to our observations, powerful extraneous stimuli caused temporary loss of time differentiation, and led to the appearance of a large number of intersignal reactions. Bromide, two grams, reduced the number of intersignal reactions, and accelerated the process of the formation of a conditioned reflex to time. Caffeine in doses of 0.5–1.0 gm. led to the appearance of a large number of intersignal reactions, and disturbed the accuracy of time reckoning.

We were unable to establish a motor-defence reflex to time to intervals of less than one minute, despite a large number of repetitions (300–500). With 30-second intervals be-

tween stimulations we observed a large number of intersignal reactions in all the animals, increasing with each successive experiment. Parallel with this, the animals became increasingly restless with each succeeding experiment, whimpering and barking and trying to bite their way out of their harness, until finally further attempts had to be abandoned.

In addition to salivary and motor conditioned reflexes to time, the conditioned reflex changes in metabolism, induced in dogs through definite time intervals, were also investigated. The work of K. M. Bykov and his colleagues on this subject demonstrated that a conditioned reflex to time could also be formed on the basis of metabolic changes. Thus, G. V. Nesterovskii and A. D. Slonim (1936) investigated thermal polypnoea in dogs (that is, accelerated respiration developed in response to a rise in the environmental temperature), having established a conditioned reflex (to the sound of the metronome) on the basis of this unconditioned reflex reaction. The authors then used the metronome every five minutes, and observed regularly recurring accesses of polypnoea at five-minute intervals, even in the absence of metronome action. R. P. Ol'nianskaia and A. D. Slonim (1938) observed the formation of conditioned reflexes to time in metabolic investigations on dogs. The animals were kept for five hours in a cold room (temperature: 10° C.) or in a warm room (temperature: 22° C.)—corresponding changes in metabolism naturally resulted—and were then returned to the kennels in which the temperature was below that of the warm, but higher than that of the cold room. Repeating these experiments systematically, the authors noted that toward the end of the five-hour stay the metabolism of dogs in the warm room began to increase, and that of dogs in the cold room, to fall.

Considering this to be the result of the formation of time associations, K. M. Bykov remarked that "these associations were established to a definite interval of time, that preceding the transfer of the dog to the kennels in which the animals' metabolism adjusted itself to the temperature of the surroundings" (1947, p. 137). This level of metabolism in the kennels was higher than in the room when heating was used in the experiments, but was lower than in the room in the cold experiments. Similar results were obtained in these experiments when the animals were kept in a room with a neutral temperature (15–16° C.) and were then transferred to a warm (22° C.) or a cold (10° C.) room. Here, the metabolism of the dog about to be transferred to a warm room fell, and that of the dog transferred to the cold room rose.

We have seen, then, that conditioned reflexes to time can be formed in dogs on the basis of quite a variety of unconditioned reactions. The investigations of conditioned time reactions in dogs have shown that these conditioned reflexes are formed with much greater difficulty than conditioned reflexes to other stimuli. They are less stable, are easily inhibited by the action of extraneous stimuli or by change in the experimental setting, and are rapidly extinguished in the absence of reinforcement. In the mechanism of their formation, conditioned time reflexes resemble trace reflexes. They are formed to trace stimulations in a variety of analysers.

These traces, recurring at equal intervals of time, form a single complex stimulus which, in combination with the basic reaction, becomes the exciter of this reaction. Like all conditioned reflexes, the conditioned reflex to time in dogs has a generalized character in the initial stage of its formation, and only later does it become more or less differentiated.

II

Conditioned reflexes to time, so thoroughly studied in dogs, have also been examined in other animals, and also in man.

P. M. Nikiforovskii (1929) investigated conditioned reflexes to time in tortoises. In one tortoise he established a conditioned motor reflex to the smell of carnation oil, with the conditioned stimulus acting every three minutes. A conditioned reflex to "pure" time was established in another tortoise, the animal being compelled to retract its head every four minutes. The author noted the development of a conditioned reflex to time in both tortoises. The first, after several experiments, began to draw in its head exactly half a minute before the exhibition of the carnation oil smell, and the second, on each occasion a short time before the expiration of the four-minute period, retracted its head, irrespective of what it was doing at the moment (in movement or at rest). The author remarked that conditioned reflexes to time in tortoises are distinguished by their extreme instability; they are extinguished by the omission of only one reinforcement, but their re-establishment requires repeated reinforcement. He also drew attention to the ease with which conditioned time reflexes in tortoises are inhibited by extraneous stimuli (e.g., by the sound of a passer's-by footsteps).

B. I. Baiandurov (1937) made a detailed study of conditioned reflexes to time in birds (pigeons). His experiments were carried out in the following manner. The electrodes of an induction apparatus, in the primary circuit of which there was a timing device closing the circuit after a definite interval, were brought into contact with the foot of a pigeon, placed in a special chamber. Movement of the foot in response to stimulation by the current was recorded by means of a lever on the smoked drum of a kymograph. Describing the results of his investigations, the author noted particularly that it was impossible to produce a conditioned reflex to short intervals of time (5–15 minutes) in birds, despite a large number of repetitions. Even the formation of a conditioned reflex to an interval of 30 minutes took place quite slowly (after 300 repetitions). The conditioned time reflex in birds was unstable and was rapidly extinguished (after two–three omissions of reinforcement). The conditioned reflex to time, when formed, was not always retained until the following day. Alcohol abolished the conditioned reflex to time. In the course of the formation of a conditioned reflex to time in birds, Baiandurov noted the appearance of involuntary movements in the intervals between the electrical stimulations (on the 25th day from the commencement of the experiment, after 150 repetitions), and these movements increased in number with each successive day, becoming more and more concentrated every time towards the time of circuit closure.

The formation of conditioned reflexes to time in various animals has been noted in a number of works by K. M. Bykov and his colleagues in connection with the analysis of dif-

ferent types of periodic changes in physiological functions. Thus, L. G. Filatova (1949), examining the daily periodicity in the long-eared hedgehog, found that in these animals the greatest muscular activity occurred during the night-time (1600–0200 hours), and activity fell to a minimum in the daylight hours. Coincident with this, there was an increase in the secretion of gastric juice in the evening and night hours. "This whole picture," wrote the author, "can be regarded as the formation of a conditioned association to time, elaborated in the process of phylogenesis in connection with the constant hunting for prey in these hours" (1949, p. 129).

By giving food to the hedgehogs in several experiments during daytime, Filatov formed new conditioned reflexes to time, as a result of which the animals manifested two periods of motor and secretory activity—one by day and one by night. From his observations the author concluded that conditioned linkage to time is a fundamental element in the daily periodicity of rest and activity in this species of animal. The formation of conditioned reactions to time in long-eared hedgehogs was confirmed by the experiments of L. A. Isaakian (1953a), who investigated the conditioned reflex thermoregulation mechanism in these animals. When the action of a certain temperature had been combined a number of times with the particular setting and time of the experiment, the author found that for several days after change of temperature the former level of metabolism was maintained. For example, when the animals were kept for 20 days in a chamber with the temperature at 22–25° C., the body temperature and oxygen consumption became established at a definite level. When the animals were transferred to a room with a temperature of 10–12° C., no noticeable change in their metabolic rate was observed during the first two–three days, an event which the author correctly regarded as the result of the formation of a conditioned reaction to the experimental setting and time of the previous experiments, and, in the absence of reinforcement, the extinction of the time association so formed took place, according to his observations, on the third to fifth day.

Natural conditioned reflexes to time, associated in their origin with the times of taking food, have been described by L. A. Isaakian (1953b) from his investigations on rabbits. His observations showed that in rabbits the temperature of the ears rises regularly during the hours of feeding. For example, the ears of rabbits which were fed at 1500 hours generally showed a raised temperature from 1200 to 1800 hours, after which the temperature fell abruptly. The same daily variations in ear temperature were seen in hungry animals, that is, in rabbits which had not been fed at the established hour. When the daily regime was changed (transfer of the feeding time from 1500 to 2000 hours), a new rhythm of ear temperature variation was formed, the maximum rise of temperature falling within the period 2000 to 0200 hours.

The importance of the development of conditioned reflexes to time in the creation of a daily periodicity in basal metabolism was noted also by A. D. Slonim (1945) and by A. G. Ponugaeva (1949) in experiments on bats.

K. B. Svechin (1952), studying changes in the daily rhythm of physiological functions in cattle, found that, in young bullocks transferred

from the conditions of maintenance in stalls to field life, a certain type of daily fluctuation in pulse rate and temperature developed, and that this was maintained on days when the external conditions (variations in the outside temperature, in relative atmospheric humidity, etc. were considerably changed from those prevailing on the preceding days. In exactly the same way also, animals transferred from field conditions to stall conditions retained for several days the periodicity in physiological functions established while they were kept in the open. The author explains the results of his investigations as indicating that a definite rhythm of changes in the environmental conditions creates a conditioned reflex to time in the animals.

The formation of conditioned reflexes to time under various conditions in monkeys is described in a number of works by K. M. Bykov and his colleagues.

The formation of the conditioned reflex to time in monkeys (Abyssinian baboon, *M. rhesus*, *M. lapunder*) was observed in an unusual form by O. P. Shcherbakova (1937, 1938, 1949). By artificially reversing the alternation of light and darkness, the author was able to reverse the animal's daily periodicity. By creating two changes of light and darkness in the 24 hours, Shcherbakova established a regime of two "nights" and two "days" in the course of 24 hours. It is a curious fact that, on transition to the previous normal light regime, the periodic changes in functions corresponding to the prolonged, artificially created periods of light and darkness, were maintained for some time. "It seemed to me completely logical to assume," wrote K. M. Bykov, in connection with these observations, "that the regular occurrence of alternation of day and night (irrespective of whether they are natural, astronomical, or created artificially for the monkeys by alternation of the periods of lighting) likewise creates a conditioned reflex to time in the cerebral cortex" (1947, p. 153).

Later K. P. Ivanov, A. R. Makarova, and A. A. Fufacheva (1953) continued these investigations of Shcherbakova on the effect of periodic alternation of light and darkness on the periodicity of physiological functions in monkeys (macaques, baboons, long-tailed monkeys). Noting that the level of gaseous exchange was lowest at the time twilight approached, these authors regarded twilight illumination as the signal stimulus, being a component of the daytime lighting and occurring also at the boundary between the two astronomical periods, day and night, which determine different conditions in the organism—the transition of the animal from daytime activity to nighttime rest. This signal stimulus enters into the natural conditioned reflex pattern. The authors note that when artificial twilight was created 30 minutes before the advent of natural twilight, the fall in metabolism at the usual time of twilight was maintained unchanged for the first two days, and only on the third day did a complete change in metabolism, corresponding to the new light conditions develop. "Thus," concluded the authors, "twilight not only causes a fall in metabolism and is of active inhibitory character, but is also closely associated with a conditioned time reflex, involving the entire daily periodicity in the physiological functions of the organism" (p. 101).

Analysing the daily periodic fluctuations in the blood sugar content of monkeys (*M. rhesus*, Abyssinian ba-

boon), E. S. Kanfor (1949) found that an increase in blood sugar occurred in relation to the feeding times, and regarded this as the result of the development of a conditioned reflex to time.

The formation of a conditioned reflex to the feeding time in monkeys was also noted by L. G. Voronin (1948). He observed the formation of a conditioned reflex to the time of feeding, manifested by the appearance of sucking movements and general motor unrest towards the times of feeding in young rhesus monkeys at about the third week.

L. G. Voronin (1951) described the formation of conditioned linkages to time in monkeys (baboons, macaques, long-tailed monkeys) in experiments in which a system of stimuli repeated at regular intervals was used. Thus, when the conditioned foodseeking reflex was repeated every two minutes, it was very soon noted that toward the end of the second minute after the administration of the stimulus the monkey ran to the feeding box and tried to reach the food. After establishment of the conditioned reflex to the stereotype of stimuli, operating at minute intervals, it could often be seen that, even in the experiment without use of the conditioned stimuli ("sham" experiment), the animal ran to the food box every minute and pressed the lever. At first the conditioned reaction to time was not absolutely accurate (the animal applied pressure 10–15 seconds before the end of the minute interval), but with continued feeding of the animal every minute, the reflex to time became progressively more and more exact.

According to the findings of Voronin and his colleagues, the conditioned reflex to "pure" time (in which only feeding of the animal was carried out at regular intervals) was more difficult to establish in monkeys. For example, in rhesus monkeys establishment of a conditioned reflex to "pure" time (1 minute) required about 400 feeding reinforcements. In Voronin's opinion, the conditioned reflex to time in these cases is formed much more quickly, but is masked by the conditioned reflex to the setting.

When the intersignal intervals are changed from one to two or three minutes, the conditioned reflex to the former time is preserved in the first two or three experiments, but fairly soon thereafter (after 6–10 experiments) a conditioned reflex to the new time is established. In monkeys it is much more difficult to establish conditioned reflexes when the intersignal intervals are prolonged to three–five minutes. When the interval between stimuli was considerably lengthened, a large number of intersignal reactions became evident, and these Voronin regarded as the manifestation of the earlier established conditioned reflexes to time and as a result of disinhibition of the conditioned reflex to the setting. The optimum interval between stimuli for the formation of a conditioned reflex to time in monkeys is thought by the author to be one–two minutes.

Voronin regarded the suppression of the conditioned reflex to the stimulus preceding the differential signal in the stereotype of stimuli, which was frequently observed in monkeys, as a special case of a conditioned reflex to time ("retrograde" inhibition of E. A. Asratian, 1934). In the experiments of Voronin and his co-workers, the system of stimuli was so arranged that every third or fourth (and sometimes sixth) stimulus was a differentiation stimulus. Consequently the second, third, or fifth stimulus was,

in Voronin's opinion, not only the signal of a positive conditioned reflex, but was also a signal for the following differentiation, so that sometimes these signals acquired also a negative signal significance. "We suggest," wrote the author, "that we are dealing here with an instance of that power of the nervous system to "reckon time," which shows itself in the form of a conditioned reflex (1951, p. 221).

O. P. Bolotina (one of Voronin's co-workers) gives a detailed description in her work (1952b) of the formation of conditioned reflexes to time in monkeys. According to her observations, such a conditioned reflex is readily formed, and begins to appear even on the first experimental day, but the time required for its stabilization increases as the intersignal interval is lengthened. The conditioned reflex forms more slowly to "pure" time than to a compound stimulus (time plus auditory or photic stimulation). In the latter case Voronin employed a pattern of sound and light stimuli, following each other at regular intervals, while reinforced and unreinforced stimuli recurred regularly in a fixed order. When the stimuli were suppressed the conditioned reflex reaction appeared at the same intervals, but in the places of the unreinforced stimulations the reflex was unstable. When the unreinforced stimulations were excluded from the pattern, the conditioned reflex to time was more stable. The conditioned reflex to "pure" time (the interval between stimuli being one minute) appeared toward the end of the third experiment, but was stabilized only after 450 repetitions.

In a special series of experiments Bolotina made observations on change in the time interval. These experiments showed that the longer the time interval was, the greater was the number of repetitions required for stabilization of the conditioned reflex to time. On transition from one interval to another, the first two days showed a conditioned reflex to the former time interval, although even from the very beginning a conditioned reflex to the new interval was also in evidence. On the following days inhibition of both the old and the new conditioned reflexes was observed, and this the author explains as the manifestation of an orientation reaction to the new arrangement of the experiment. From the third day intersignal reactions appeared in large number, but these were gradually controlled and became more and more concentrated toward the end of the new time interval. The rapid formation of conditioned reflexes to time in monkeys and their facile transition from one time interval to another are explained, according to the author, by the great mobility of the nervous processes in these animals.

That the conditioned reflex to time in monkeys is more stable than in other animals was indicated by the absence of any appreciable inhibitory effect from the operation of an extraneous stimulus (whistle) towards the end of the intersignal interval. During extinction of the conditioned reflex to time a large number of intersignal reactions first appeared, and only later (after from 1 hr. 33 min. to 2 hr. 6 min.) did complete extinction develop. Extinction took place in a wave-like manner. The extinguished reflex was easily re-established. In monkeys the conditioned reflex to time was disturbed by any pronounced change in the experimental setting, such as, for example, delay of the start of the experiment by 3–10 minutes.

In her own subsequent investiga-

tions O. P. Bolotina (1953) found that small doses of bromide (0.2–0.35 gm.) had a certain controlling effect on conditioned reflex activity, evidenced by a reduction in the number of intersignal reactions. Larger doses (2–3 gm.) exercised a sustained positive action on the reflex to time. A mixture of bromide and caffeine had a more definite, positive effect. Caffeine, in doses of 0.02–0.07 gm., produced an increase in the number of intersignal reactions in all the monkeys.

According to the findings of O. P. Bolotina and A. A. Popova (1953), phenamine, in a dose of 0.3 mg./kg., led to slight disturbance of time reckoning in the monkeys in the daytime, with increase in the number of intersignal reactions, whereas during nighttime it rendered the conditioned reflex to time more exact. In a dose of 1.5 mg./kg., it caused considerable inhibition of reflexes to time during the day, but in the hours of night it intensified the reflexes and made them more exact.

In the view of Voronin and his coworkers, the conditioned reflex to time in monkeys " . . . is formed on trace stimulations coming from the motor analyser (pressure on the lever) and on trace stimulations from the digestive tract, these being constantly combined in relation to identical segments of time. These traces in the cerebral cortex are built up into a definite physiological state of the cortical cells. This state becomes a conditioned stimulus for the nervous system which adjusts the entire organism to meet the material conditions of the surroundings" (Bolotina, 1952b, p. 203).

Thus, the formation of conditioned reflexes to time has been established for animals of quite a variety of classes (reptiles, birds, mammals), and among mammals, for representatives of various divisions (insectivores, cheiropterids, rodents, carnivores, artiodactyls, monkeys). If we may judge from the similarity of the process of formation of conditioned reflexes to time in the different animals (the relative difficulty of their production in comparison with conditioned reflexes to other stimuli, their facile inhibition, their rapid extinction in the absence of reinforcement, etc.), the nervous mechanism involved in the process is basically the same in all vertebrates. In the natural living conditions of many animals the formation of conditioned reflexes to time is of considerable importance in the development of a periodicity of physiological functions, typical for the particular animals. Investigations on the conditioned reflexes to time in various animals confirm Pavlov's view on the mechanism for time reckoning by the nervous system, and confirm that the power to reckon time is inherent in the cortical cells of all analysers, and that it is unnecessary to postulate the existence of a special cortical "time analyser," as has been suggested by several physiologists (P. M. Nikiforovskii and others).

The results of investigations on conditioned reflexes to time in animals may find wide application in practical animal husbandry, in relation to the development and stabilization of rational regimes for the maintenance and economic exploitation of farm animals. In relation to periodically recurring processes (feeding, watering, moving cattle, driving out to pasture, milking, rest) the establishment of conditioned reflexes to the times of these processes, that is, the establishment of a complex time pattern, is important, among other measures, in the management of animals. The importance attaching to the development of such a pattern is

determined, not only by the fact that it instils order into and facilitates the work of man in caring for animals, but also because the positive physiological effect of the measures being pursued is considerably enhanced thereby. Time, as a unique exciter of reflex activity, is capable (in conjunction with other stimuli) of creating and strengthening a definite course in physiological processes, of giving them a definite rhythm, a definite interrelationship in time. This undoubtedly creates more favorable conditions for the course of each of the processes. It is for this reason that the observation of strict regularity in the feeding times of animals increases the productivity of the food, that the strict maintenance of a regular milking routine increases the yield, and so on.

In State farm and collective farm practice measures are taken to create a constant regime of animal maintenance: times and routines for feeding are laid down (this is often done separately for different groups of animals), and definite times are fixed for exercising the animals, for milking cows, etc. Numerous examples of these measures are given in the work of A. V. Kvasnitskii and V. A. Koniukhova (1954). In agreement with the results of laboratory investigations on conditioned reflexes to time, their formation and consolidation in practical animal farming occurs more rapidly when the conditioned reflex is established not to "pure" time, but to a complex stimulus. In practical work also an attempt is made to combine the time of a particular productive process with some kinds of concomitant signals. In some cases, for example, feeding time was linked with the rapid striking of a rail, exercise time with slow striking, and so on. Very often such concomitant signals, strengthening the conditioned reflex to time that had been formed, were verbal stimuli (being, of course, only sound stimuli to the animals). Thus, the arrival of exercise time was linked with the word "walk," the arrival of the milking time for each cow with the pronouncing of its name, etc.

Change in the established regime, that is, a change from some earlier established conditioned time reflexes to others, presents a certain amount of difficulty for the nervous system, as can be judged from the results of laboratory investigations. The same has been observed in animal husbandry under natural conditions. A. V. Kvasnitskii and V. A. Koniukhova (1954) illustrate this by the following example. Special investigations showed that in pigs under the usual feeding conditions (morning, noon and evening) the secretion of gastric juice is maximum towards the feeding times, and falls during the rest of the time, particularly during the night. When the feeding times were changed (evening, midnight and early morning) the previous conditioned reflex to time, that is, increase in the secretion of gastric juice at the times of the previous day's feeding, was preserved for several days, and only gradually was a reflex to the new feeding times established. Such a reconstruction of the time pattern entailed a period of disturbance in the functioning of the coordinated organs, and indeed any disturbance of the established regime produced the same result. Kvasnitskii and Koniukhova cite as an example the reduction in the milk-yield by five per cent and in the fat content of the milk to 0.2–0.4 per cent that resulted from milking 30–40 minutes later than the usual time.

While we stress the importance of

the constant maintenance of a set daily routine, it should at the same time be noted that the most rational regime varies for different conditions. Consequently, in the course of animal-rearing practice the question of a change of regime arises very frequently (in connection with the transfer of the animal from one production group to another, with change in season etc.). It is, of course, obvious that in these circumstances the disruption of the old time pattern will be accompanied by a drop in the animals' productivity on the first few occasions. In order, therefore, to facilitate this readjustment, this adaptation of the animals to the new conditions, it is essential that one should understand the laws governing the formation and reconstruction of conditioned reflexes to time in animals.

In man the physiological mechanism involved in the reckoning of time by the nervous system has not been adequately studied. The fact itself of the perception of time by man has, of course, been known for a long time. The perception of time by man and the various changes that occur in this perception under both normal and pathological conditions have long been the object of investigation by many psychologists, psychiatrists and neurologists. It was, however, only the doctrine developed by Pavlov on higher nervous activity that made it possible to link the investigation of such matters with physiological investigations on the functions of the cerebral cortex, with investigations on conditioned reflex activity, and particularly conditioned reflexes to time in man.

A number of works by Iu. P. Frolov (1924, 1928, 1951) have been devoted to an analysis of conditioned time reflexes in man. Believing, in agreement with the views of Pavlov, that " . . . dynamic changes in nerve traces, as the 'differentials' in the reckoning, constitute the fundamental element in the reckoning of time" (1951, p. 837), Frolov pointed out that the speed with which traces from stimuli are extinguished may vary with the state of the cortical cells, and that various errors in the assessment of time are, in his opinion, connected therewith.

According to his findings, a tendency to over-assess time (within limits of five–eight per cent) was constantly observed in experiments on healthy individuals of various ages and occupations. The author explained this by the development of internal inhibition in the subjects under the experimental conditions (darkness, silence), in consequence of which the trace from the signal-stimulus, from which the reckoning of time began, was damped down more rapidly than usual. In other words, trace processes reached a certain degree of intensity earlier than they should, and therefore the true interval of time seemed longer to the subject ("time was extended"). The same was observed, according to the author, in the case of anticipation. When the excitability of the cortical cells was heightened, on the other hand, time was under-assessed.

Frolov described pronounced disturbances of time reckoning in various psychical illnesses (over-assessment of time in depression, Korsakov's psychosis, etc.), and associated these disturbances with the development of an inhibitory process in the cerebral cortex, and the more rapid suppression of traces from the stimuli.

"From a practical point of view it is important to note," the author concluded, "that all changes in the state

of the subjects' health, the degree of their adaptation to work, and also the presence of emotional elation or depression were immediately reflected in the accuracy of their time reckoning, inducing acceleration or slowing in the calculation, over-assessment or under-assessment of the passage of time" (1951, p. 838).

During the establishment of a rhythmical pattern, the formation of conditioned reflexes to time in man was noted and examined in connection particularly with the analysis of one of the phenomena observed during the process, and first described by V. M. Bekhterev (1908).

Examining the motor reactions of man to rhythmically recurring sounds, Bekhterev observed that, after sudden cessation of the sound, the subject continued to execute several movements at the former rate. Also, as the period over which the sounds were in operation increased, and as the rate at which they followed one another was raised, so the number of motor reactions executed after the cessation of the sound increased. Subsequently, G. P. Zelenyi (1923) made a special study of this phenomenon by setting the subject to beat in time with a metronome (120 beats per minute). When the metronome was stopped (after 120 beats) the subject made several additional movements at the same rate, and this occurred even when the subject was asked not to make any excess movements. The author also observed that when, after an interruption in the rhythmical work, the metronome gave one beat, the subject made two or several movements in succession with the same intervals between them as between the beats of the metronome. In a joint work with B. N. Kadykov, G. P. Zelenyi (1937) found that the motor reactions very often outstripped the conditioned signals, and developed before the next succeeding beat of the metronome. According to the observations of K. M. Bykov (1925), the rhythm of the movements does not always coincide with the rhythm of the metronome; some subjects offer slower movements, some more rapid, and some again make movements at an irregular rate, while in only 49 per cent of the subjects was the rate of the movements coincident with the beats of the metronome. According to Bykov's findings, the number of excess movements after cessation of the metronome did not exceed three in the majority of subjects (76 per cent), although in some cases it reached nine. Both Bykov and Zelinyi regard these "excess" movements as the result of the development of a conditioned reaction to time.

Detailed investigations of the same phenomenon were carried out recently by M. A. Alekseev (1953) with various rates of movements (from 20–30 to 200 per minute, that is with intervals between successive conditioned signals of from 0.3 to two–three seconds). His observations showed that supplementary motor reactions after cessation of the rhythmical signals only developed when the intervals between the conditioned signals were less than one second. Regarding these reactions as a manifestation of a conditioned reaction to time, the author also points to other forms of this phenomenon (shortening of the latent period of the motor reactions, frequent anticipation of the next signal). With intervals exceeding one second (1.6–3 seconds) no supplementary movements appeared as a rule, that is, the conditioned reaction to time was inhibited. In these cases a sharp fall in cortical excitability was seen immediately

after the conditioned signal, to be replaced by heightened excitability as the moment for the action of the next signal approached. This the author regarded as a "latent" manifestation of the conditioned reaction to time. Assuming that the development of a conditioned reaction to time is determined by the degree of concentration of the conditioned excitation in the period leading up to the next reinforcement, the author explains the differences in the formation of this reaction at different intervals as follows: in the presence of large intervals following negative induction, excitability increases slowly, while in the presence of short intervals it rises steeply. Since between the conditioned signal (beat of the metronome) and the complex stimulus which induces the conditioned reaction to time, there occur, according to the author, active induction relationships, it follows that in the case of longer intervals (as a result of the slow increase in excitability), the external signal proves stronger, and the reaction to time is inhibited. Whereas with short intervals (in consequence of the steeper rise in cortical excitability) " . . . this state of the cortical cells emerges as an independent conditioned stimulus, inhibiting the reaction to the sound and inducing the motor reaction" (Alekseev, 1953, p. 896).

The formation of conditioned reflexes to time has been noted also in the course of investigations on the daily periodicity in the physiological functions of man. Thus, investigations of daily variations in the composition of the blood in man have shown that in many people the number of leucocytes increases at the usual hours of eating (Voronov & Riskin, 1925; Orlova, 1937), and that this increase is the result of the formation of a conditioned reflex to the time of taking food (Belen'kii, 1949). A. G. Urin and E. S. Zenkevich (1952) stated that not less than six–seven days were required for the formation of a conditioned leucocytic reflex to the time of taking food. When the time of eating was changed, the conditioned leucocytic reflex was extinguished after two–four days, and the longer the conditioned reflex had taken to become established, the slower was its extinction. The extinguished reflex was restored rapidly on return to the previous conditions of eating.

The features attaching to the formation of a conditioned reaction to time in man were investigated in our laboratory by A. S. Dmitriev. The observations were made on children aged 8 to 14 years. Conditioned motor reactions were established in the subjects to sounds with verbal reinforcement, the conditioned stimulus being repeated at equal intervals of time (25–30 seconds).

In many of the children the formation of a conditioned reaction to time began with the appearance of intersignal motor reactions (after 5–13 repetitions), and indeed these reactions appeared also at times later (every five–six repetitions or even more frequently) at quite varied times in the intersignal interval. The character of the distribution of all the intersignal reactions in one of the subjects in the first experiment is shown by the dotted line curve in Fig. 3. In subsequent experiments, however, the number of intersignal reactions diminished, and appeared mainly in the second half of the intersignal interval (as is shown by the continuous line curve in Fig. 3). Finally, a more or less differentiated conditioned reaction to time was formed, and the intersignal reactions

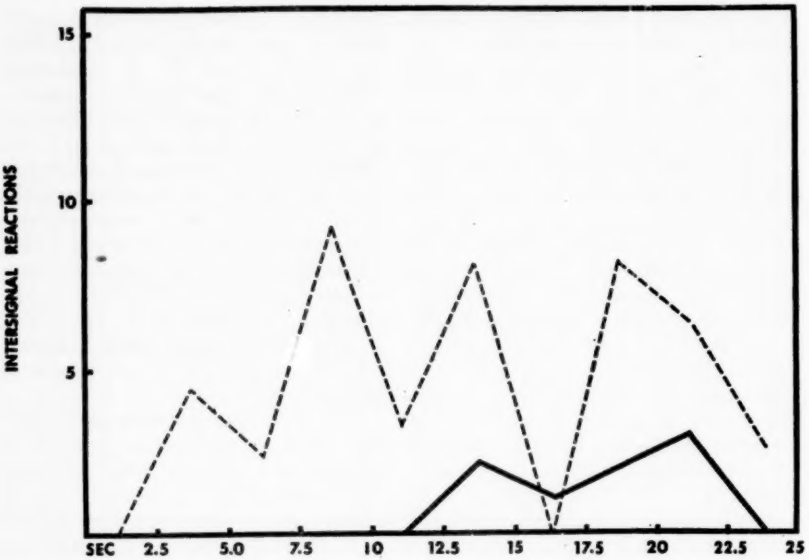

FIG. 3. DETAILS OF THE APPEARANCE OF INTERSIGNAL MOTOR REACTIONS IN VARIOUS STAGES OF FORMATION OF THE MOTOR REACTION TO TIME IN CHILDREN (SUBJECT T.K.).
The dotted line shows the distribution of the total number of intersignal reactions in the 25-second intersignal interval seen in experiment No. 1, and the solid line shows the distribution of the total number of intersignal reactions in experiment No. 2.

appeared two–three seconds before the next conditioned signal; when the next signal was omitted, reactions often developed at the time when it was due. Thus, the process of formation of a conditioned reaction to time passes through the same stages in children as we observed during the formation of a motor-defence conditioned reflex to time in dogs (cf. Fig. 1 and 3). The duration of the stages and their relationships to one another differed in different children, as the graphical illustrations in Fig. 4 show. The total time required for the formation of a differentiated conditioned reaction to time in children varied from 29 to 82 repetitions.

It must be noted, however, that the formation of a conditioned reaction to time in this form was noted predominantly in children of early school age. But some of these intersignal reactions appeared in small numbers even after a large number of repetitions, a fact which prevented exact observation of all the stages in the formation of a conditioned reflex to time. In children of middle school age the conditioned reaction to time appeared in the form of intersignal reactions in only 42.8 per cent of all cases; in most no intersignal reactions whatever were seen. It was, nevertheless, possible to establish a conditioned linkage to time in this group of children, but by a different method: being told (after four–five experiments) to make independent pressures at the same intervals as in

the preceding experiments, they preserved with a fair degree of accuracy the usual intersignal interval between the pressures, which was possible only in the presence of a more or less exact reflection of this interval in the child's cerebral cortex. The older the children, the greater was the accuracy with which this task was carried out.

These investigations enabled us to conclude that, in the case of children of early school age, first signal system associations occupy a prominent place in the formation of the conditioned reaction to time, while in the case of children of middle school age second signal system connections play the leading part in the formation of the conditioned time reaction, inhibiting through negative induction the appearance of the conditioned reaction within the limits of the first signal system in the form of intersignal motor reactions. This feature in the formation of conditioned reactions to time should, obviously, be more pronounced in adults.

There is no doubt that investigations of conditioned reactions to time in man are important for the solution of a number of practical problems. Such problems include the establishment and stabilization of a rational regime of work and rest, the arrangement of the school-day for children, the rhythm of industrial processes, nutritional regimes, etc. Investigations of conditioned reactions to time in man will lead to discovery of the physiological mechanism involved in

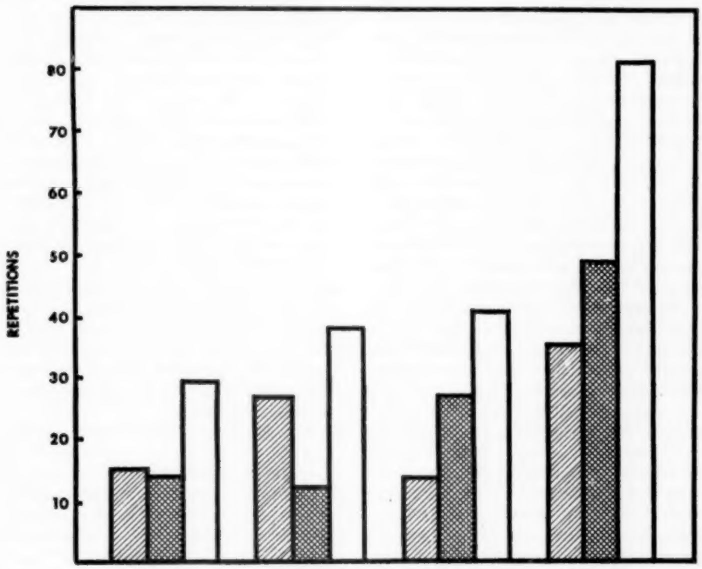

FIG. 4. DURATION OF THE STAGES IN THE FORMATION OF A CONDITIONED MOTOR REACTION TO TIME IN CHILDREN.
Notation as in Fig. 2.

the perception and measurement of time and its features in individuals of different ages, and such information might be used to advantage in training and education. Thus, further study of the conditioned reactions to time in man is closely bound up with the solution of important practical problems in medicine, psychology, and education.

REFERENCES

ALEKSEEV, M. A. K voprosu o nervnykh mekhanizmakh i vzaimodeistvii dvukh korkovykh signal'nykh sistem pri ritmicheskikh dvigatel'nykh uslovnykh reaktsiiakh cheloveka. [Nervous mechanisms and interactions between the two cortical signal systems in rhythmical motor conditioned reactions in man.] *Zhurn. vyssh. nervn. deiat.*, 1953, **3**, 883–897.

ASRATIAN, E. A. K fiziologii irradiatsii i kontsentratsii protsessov v kore bol'shikh polusharii. [Physiology of irradiation and concentration of processes in the cerebral cortex.] *Dokl. Akad. nauk*, 1934, **4**, No. 3; Sistemnost' raboty bol'shikh polusharii golovnogo mozga. [Systematization in the functioning of the cerebral cortex.] *Tr. fiziol. lab. im. akad. I. P. Pavlova.* [Proc. Pavlov Physiol. Lab.] 1938, **8**.

BAIANDUROV, B. I. *Uslovnyi refleksy u ptits* [*Conditioned reflexes in birds.*] Tomsk, 1937.

BEKHTEREV, V. M. O reproduktivnoi i sochetatel'noi reaktsii v dvizheniiakh. [Reproductive and combining reactions in movements.] *Obozrenie psikhiatrii, nevrologii i eksper. psikhologii*, 1908, No. 7, 385–389.

BELEN'KII, G. S. Ob uslovnom pishchevom leikotsitoze (Conditioned alimentary leucocytosis) Communication No. 1 in Symposium: *Nervnogumoral'naia reguliatsiia deiatel'nosti pishchevaritel'nogo apparata.* [Neurohumoral regulation of the activity of the digestive apparatus.] 1949, 265–280.

BERITOV, I. S. *Individual'no priobretennaia deiatel'nost' tsentral'noi nervnoi sistemy.* [Individually acquired activity of the central nervous system.] Tbilisi, 1932.

BOLOTINA, O. P. Dvigatel'nye uslovnye refleksy na vremia u sobak. [Motor conditioned reflexes to time in dogs.] *Tr. In-ta fiziol. im. I. P. Pavlova.* [Proc. Pavlov Physiol. Inst.], 1952a, **1**, 29–34; Dvigatel'nye uslovnye refleksy na vremia u obez'ian. [Motor conditioned reflexes to time in monkeys.] Ibid., 1952b, **1**, 196–204; Vliianie broma i kofeina na uslovnye refleksy na vremia u sobak i obez'ian. [Effect of bromide and caffeine on conditioned reflexes to time in dogs and monkeys.] Ibid., 1953, **2**, 52–63.

BOLOTINA, O. P., & POPOVA, A. A. Vliianie fenamina na dvigatel'nye uslovnye refleksy na vremia u nizshikh obez'ian. [Effect of phenamine on motor conditioned reflexes to time in the lower Apes.] *Tr. In-ta fiziol. im. I. P. Pavlova.* [Proc. Pavlov Physiol. Inst.], 1953, **2**, 64–68.

BYKOV, K. M. O skorosti tormoznoi reaktsii u cheloveka. [Speed of the Inhibiting Reaction in Man.] In Symposium: *Opyt nauchnogo izucheniia voprosov fizicheskoi kul'tury.* [An attempt at the scientific investigation of physical training problems.] Leningrad, 1925; *Kora golovnogo mozga i vnutrennie organy.* [Cerebral cortex and the internal organs.] Medgiz, 1947.

BYKOV, K. M., & SLONIM, A. D. Kortikal'nye mekhanizmy fiziologii vremeni v organizme zhivnotykh i cheloveka. [Cortical mechanisms in the physiology of time in the organism of animals and man.] In Symposium: *Opyt izucheniia period. izmenenii fiziol. funktsii v organizme.* [An attempt at the study of periodic changes in the physiological functions of the body.] 1949, 3–14.

DERIABIN, V. S. Dal'neishie materialy k fiziolofii vremeni, kak uslovnogo vozbuditelia sliunnykh zhelez. [Further material on the physiology of time, as a conditioned stimulus of the salivary glands.] Diss., St. Petersburg, 1916.

DOBROVOL'SKII, V. M. O pishchevykh sledovykh refleksakh. [Alimentary trace reflexes.] Diss., St. Petersburg, 1911.

FEOKRITOVA, IU. P. *Vremia kak uslovnyi vozbuditel' sliunnoi zhelezy.* [Time as a conditioned excitant of the salivary gland.] St. Petersburg, 1912.

FILATOVA, L. G. Sutochnyi ritm u gryzunov i nasekomoiadnykh i opyt ego eksperimental'nogo izucheniia. [Daily rhythm in rodents and insectivorous animals, and an attempt at their experimental investigation.] In Symposium: *Opyt izucheniia period. izmenenii fiziolog. funktsii v organizme.* [Attempt at the investigation of the periodic changes in the physiological functions of the body.] 1949, 116–130.

FROLOV, IU. P. K fiziologii tak nazyvaemogo "chuvstva vremeni." [Physiology of the "Time Sense."] Paper delivered at Second

Congress of Psychoneurologists, 1924; Fiziologicheskii mekhanizm otscheta vremeni v nervnoi sisteme zhivotnykh i cheloveka. [Physiological mechanism for the reckoning of time in the nervous system of animals and man.] Nauchnoe slovo, 1928, Nos. 6–7; Fiziologicheskoe uchenie I. P. Pavlova o vremeni kak svoeobraznom razdrazhitele nervnoi sistemy. [Pavlov's doctrine of time as an independent stimulus of the nervous system.] *Zhurn. vyssh. nervn. deiat.*, 1951, **1**, 831–839.

GAMBARIAN, L. S. K vosprosy ob sulovnykh oboronitel'nykh refleksakh. [Conditioned defence reflexes.] *Tr. In-ta fiziol. im. I. P. Pavlova*, 1952, **1**, 73–84.

GROSSMAN, F. S. Materialy k fiziologii sledovykh uslovnykh sliunnykh refleksov. [Contribution to the physiology of trace conditioned salivary reflexes.] Diss. St. Petersburg, 1909.

ISAAKIAN, L. A. *Ob uslovnoreflektornom mekhanizme termoreguliatsii u iezhei.* [The conditioned reflex mechanism for thermoregulation in hedgehogs.] In Symposium: *Opyt izucheniia reguliatsil fiziol. funktsii.* [An attempt at the investigation of the regulation of physiological functions.] 1953a, **2**, 46–54; *O'vzaimo otmoshenii sosudistykh reaktsii i gazoobmena.* [Interrelationships between vascular reactions and gas exchange.] Ibid., 1953b, **2**, 55–63.

IVANOV, K. P., MAKAROVA, A. R., & FUFACHEVA, A. A., Vliianie intensivnosti osveschenija na gazoobmen u obez'ian. [Effect of the intensity of lighting on gas exchange in monkeys.] In Symposium: *Opyt izucheniia reguliatsii fiziol. funktsii.* [An attempt at the investigation of the regulation of physiological functions.] **2**, 94–101.

KANFOR, E. S. Vliianie fenamina na sutochnyi ritm fiziologicheskikh protsessov u obez'ian. [Effect of phenamine on the daily rhythm of physiological processes in monkeys.] In Symposium: *Opyt izucheniia period. izmenenii fiziolog. funktsii v organisme.* [An attempt at investigation of periodic changes in the physiological functions of the body.] 1949, 87–95.

KLESHCHOV, S. V. Vliianie nebol'shikh izmenenii promezhutkov vremeni mezhdu razdrazhiteliami na velichinu uslovnogo refleksa. [Effect of slight changes in the time interval between stimuli on the magnitude of a conditioned reflex.] *Tr. fiziol. labor. im akad. I. P. Pavlova*, 1941, **10**, 255.

KRZHISHKOVSKII, K. N. K fiziologii uslovnogo tormoza. [On the physiology of conditioned inhibition.] *Tr. Ob-va russkikh vrachei v S.Pb.* [Proc. Russian Medical Society of St. Petersburg.] 1908.

KUPALOV, P. S. O mekhanizme vzaimodeistviia tormoznykh i aktivnykh punktov v kore bol'shikh polysharii pri funktsional'noi mozaike. [Mechanism involved in the interaction of inhibitory and active points in the cerebral cortex as part of the functional mosaic.] *Tr. fiziol. labor. im. akad. I. P. Pavlova*, 1929, **3**, Nos. 2–3; O sostoianii kory bol'shikh polusharii v intervalakh mezhdu primeneniem uslovnykh razdrazhitelei. [The state of the cerebral cortex in the intervals between applications of the conditioned stimuli.] *Arkh. biol. nauk*, 1931, **31**, 4; O vliianii sistemy ritmicheskikh uslovnykh refleksov na obrazovanie nobogo uslovnogo refleksa. [Effect of a system of rhythmical conditioned reflexes on the formation of a new conditioned reflex.] *Tr. fiziol. labor. im. I. P. Pavlova*, 1933, **5**.

KVASNITSKII, A. V., & KONIUKHOVA, V. A. Primenenie ucheniia I. P. Pavlova v zhivotnovodstve. [Application of Pavlov doctrine in animal husbandry.] *Izd. Akad. nauk UkSSR, Kiev*, 1954.

MAIOROV, F. P. Maibolee slozhnye fakty iz fiziologii vysshei nervnoi deiatel'nosti. [The most complex aspects of higher nervous activity.] *Tr. fiziol. labor. im. I. P. Pavlova*, 1933, **5**, 255.

NESTEROVSKII, G. V., & SLONIM, A. D. *Uslovnoreflektornyi mekhanizm teplovoi odyshki.* [Conditioned reflex mechanism of heat asthma.] *Biull. V.I.E.M.* [*All-Union Institute of Exp. Med.*] 1936, 3–4.

NIKIFOROVSKII, P. M. K fiziologii vremeni. [Physiology of time.] *Russkii fiziol. zhurn.* 1929, **12**, 483–490; Materialy k ucheniiu o vyssei nervnoi deiatel'nosti. [Contribution to the study of higher nervous activity.] *Zhurn. vyssh. nervn. deiat.*, 1951, **1**, 827–830.

OL'NIANSKAIA, R. P., & SLONIM, A. D. Vliianie kory golovnogo mozga na reguliatsiiu tepla v organizme. [Influence of the cortex on heat regulation in the body.] Soobshch. I [Paper No. 1], *Fiziol. zhurn. SSSR*, 1938, **25**, No. 6.

ORLOVA, S. D. *Arkh. biol. nauk*, 1937, **44**, No. 3, 120.

PAVLOV, I. P. *Lektsii o rabote bol'shikh polusharii golovnogo mozga.* [Lectures on the functioning of the cerebral hemispheres.] *Complete Works*, 1947a, **4**; *Dvadtsatiletnii opyt ob"ektivnogo izucheniia vysshei nervnoi deiatel'nosti.* [Twenty years' experience in the objective study of higher nervous activity.] *Complete Works*, 1947b, **3**.

PIMENOV, P. P. *Osobaia gruppa uslovnykh refleksov.* [A special group of conditioned reflexes.] Diss., St. Petersburg, 1907.

PONUGAEVA, A. G. Stadnaia forma sushchestvovaniia i sutochnaia periodika u letuchikh myshei. [Gregariousness and daily periodicity in bats.] In Symposium: *Opyt izucheniia period. izmenenii fiziol. funktsii v organizme.* [Periodic changes in the physiological functions of the body.] 1949, 104–115.

SHCHERBAKOVA, O. P. Sutochnaia periodika fiziologicheskikh funktsii u mlekopitaiushchikh. [Daily periodicity in the physiological functions of mammals.] *Biull. eksper. biol. i med.*, 1937, **4**, 335; Eksperimental'noe izuchenie sutochnoi periodiki u zhivotnykh. [Experimental investigation on daily periodicity in animals.] Ibid., 1938, **5**, 167; Eksperimental'noe izuchenie sutochnogo ritma fiziologicheskikh funktsii u obez'ian. [Experimental investigation on the daily rhythm in the physiological functions of monkeys.] In Symposium: *Opyt izucheniia period. izmenenii fiziol. funktsii v organizme.* [Investigation of periodic changes in the physiological functions of the body.] 1949, 42–64.

SLONIM, A. D. *Izv. Akad. nauk SSSR, biol. series*, 1945, No. 3.

STUKOVA, M. M. Dal'neishie materialy k fiziologii vremeni kak uslovnogo vozbuditelia sliunnykh zhelez. [Further contribution on the physiology of time as a conditioned excitant of the salivary glands.] Diss., St. Petersburg, 1914.

SVECHIN, K. B. Sutochnyi ritm fiziologicheskikh funktsii u krupnogo rogatogo skota v usloviiakh lagernogo i stoilovogo soderzhaniia. [Daily rhythm of physiological functions in cattle kept under field and stall conditions.] *Fiziol. zhurn. SSSR*, 1952, **38**, 319–325.

URIN, A. G., & ZENKEVICH, E. S. Pishchevoi leikotsitoz i uslovnyi leikotsitarnyi refleks na vremia priema pishchi. [Alimentary leucocytosis and the conditioned leucocytic reflex to the time of taking food.] *Zhurn. vyssh. nervn. deiat.*, 1952, **2**, 715–722.

VASILENKO, F. D. K voprosu ob uslovnom reflekse na vremia. [The problem of the conditioned reflex to time.] *Tr. fiziol. lab. im. akad. I. P. Pavlova.* [Proc. Pavlov Physiol. Lab.] 1932, **4**, 310.

VATSURO, E. G. Refleks na vremia v sisteme uslovnykh razdrazhitelei. [Reflex to time in a system of conditioned stimuli.] Ibid., 1948, **13**, 5–20.

VORONIN, L. G. K voprosu o razvitii bezuslovnykh i uslovnykh refleksov u novorozhdennykh detenyshei makakov rezusov. [The development of unconditioned and conditioned reflexes in the newborn young of *M. rhesus.*] *Fiziolog. zhurn. SSSR*, 1948, **334**, 333–338; Nekotorye dannye ob uslovnom reflekse na vremia u nizshikh obez'ian (pavianov, makakov i martyshek). [Some findings on the conditioned reflex to time in the lower apes (baboons, macaques, long-tailed monkeys).] *Zhurn. vyssh. nervn. deiat.*, 1951, No. 1, 213–222.

VORONOV, A., & RISKIN, I. *Russkaia klinika*, 1925, **3**, No. 12, 483.

ZAVADSKII, I. V. Materialy k voprosu o tormozhenii i rastormazhivanii uslovnykh refleksov. [Contribution to the inhibition and disinhibition of conditioned reflexes.] Diss., St. Petersburg, 1908.

ZELENYI, G. P. Materialy k voprosu o reaktsii sobaki na zvukovye razdrazheniia. [Reactions of the dog to auditory stimuli.] Diss., St. Petersburg, 1907; O ritmicheskikh myshechnykh dvizheniiakh. [Rhythmical muscle movements.] *Russkii fiziol. zhurn.*, 1923, **6**, 155.

ZELENYI, G. P., & KADYKOV, B. N. In Symposium: *Problemy motoriki v nevrologii i psikhiatrii.* [Motor functions in neurology and psychiatry.] 1937, 84.

Received July 16, 1958.

PSYCHOLOGCIAL BULLETIN
VOL. 56, NO. 2, 1959

THE EFFECTS OF HIGH INTENSITY INTERMITTENT SOUND ON PERFORMANCE, FEELING, AND PHYSIOLOGY[1]

ROBERT PLUTCHIK
Hofstra College

There is one aspect of the general problem of the effect of high intensity noise on man that has been insufficiently explored; this concerns the question of the effects of irregular or interrupted sound on performance, feeling, and physiology. Evidence which has been accumulating during the past few years suggests that such intermittent sound has both different and more disturbing effects on *S*s than those of steady sound sources. Such differences may, in addition, have important theoretical implications for the understanding of brain mechanisms.

This report will, therefore, attempt a review and analysis of the literature dealing with the problem of the effects of interrupted sound on various measures of human functioning, particularly as related to perceptual-motor activity, subjective report, and body physiology.

In 1946, Berrien published a review of the effects of noise as applied mainly to industrial environments and reported that, although much of the evidence was inconclusive, there were some indications that noise tended to affect work output and speed of work. This review noted that there are marked individual differences in susceptibility to the ill-effects of noise. No attempt to distinguish the effects of intermittent from steady noise was made except to note the increased unpleasantness of intermittent sounds.

Kryter in 1950 published an extensive review of the literature of the effects of noise on men. His conclusions from studies conducted under laboratory conditions indicated that experimental studies can be grouped under three categories:

1. Experiments which demonstrated deleterious effects of noise. Nearly all, if not all, can be heavily criticized on one or more points so that findings can be accepted only with considerable reservations.
2. Experiments which demonstrated slight, inconsistent or inconclusive detrimental effects from noise.
3. Experiments that demonstrated that man can do certain types of muscular and mental work as efficiently and productively in noise as in quiet, even for prolonged periods. For some few tasks, noise apparently improved performance.

However, it is obvious that previous studies have not sampled all the different types of behaviors of which man is capable. Kryter also notes that different kinds of sounds may have marked effects on feelings as well as on certain physiological measures. These will be described in later sections of this review.

Since Kryter's report, a number of studies have appeared which indicate that noise can affect performance.

An experiment by Miller (1953) tested the effects of a 90 db,[2] 8000 cps tone on the four measures of critical flicker fusion, cancelling c's, word

[1] Supported by Contract Nonr-2252(01) with the Office of Naval Research. Reproduction for any purpose of the United States Government is permitted.

[2] All designations of intensity in this report will be decibels with reference to 0.0002 dynes/cm. sq. = 0 db unless otherwise noted.

fluency, and trembling. Only the last measure showed any effect (i.e., increase in trembling) as a result of the sound. In the Benox report, Miles (1953) indicated that *S*s exposed to a 115 db sound for three hour periods showed an impaired performance on Coordinated Serial Reaction Time—i.e., on a task in which the *S* directed a beam of light at a series of targets by means of airplane controls. This was reported as being the most effective of all the tests for showing an effect of noise on performance.

Broadbent (1951a) described an experiment in which the *S* was required to observe and correct any pointer that exceeded the danger mark on a series of steam pressure gauges. The pointer signals were presented at random in a series of test runs each 90 minutes long. Ten *S*s participated, each for five days, and were exposed in the sequence: quiet, quiet, noise, noise, quiet. Noise levels for quiet were 70 db, and noise 100 db. The results indicated that *S*s performed more poorly on the noise days than on the quiet days preceding and following the noise sessions. When the task was made simpler (the dials were replaced by lights whose brilliance was sharply intensified at random as a signal) another group of *S*s showed no significant effect of noise.

Other studies by Broadbent (1951b, 1953, 1954) were designed to investigate the effects of pacing on the performance of a vigilance task under continuous 100 db noise conditions. More errors were made on the task under noise conditions than under control conditions in each experiment. However, the results of the pacing were inconsistent in two studies. In one study it was found that fewer errors were made by the group under unpaced conditions (stimulus presented after every response) than by the group under paced conditions (stimulus presented every second). In the other study no difference was found between paced and unpaced performance. Unlike the first study, however, in the latter task, performance was prolonged over a one and one-half hour period suggesting that the *S*s may have adapted to the pacing and noise conditions. More errors were made under the noise condition.

Using a modification of the above procedure, Jerison and Wing (1957) had *S*s watch three clocks side by side whose pointers made double jumps on a random basis of about one per minute. The *S*s watched the clocks during two hour sessions and pressed a key when each double jump occurred. Comparing a "quiet" condition of 83 db with a noise condition of 114 db produced by a loudspeaker with frequencies of 20 to 9600 cps, they found that performance decreased significantly in the final half-hour of the noise condition. No change in performance occurred during the two-hour "quiet" condition.

In another clock-watching experiment Jerison (1956) reported a definite impairment of performance with increased time in the task situation, and concluded that this was due partly to noise, partly to fatigue, and partly to the unreliability (test-retest) of the task itself. An important, but unexpected observation was that *S*s who showed marked auditory fatigue after the test tended to perform at the same level in noise as in quiet, whereas those *S*s with only mild hearing losses tended to fall off from their starting performance level. This finding emphasizes the important fact of individual differences.

Another effect of noise that has very recently been discovered is the fact that time estimation is affected

by noise levels. Hirsh, Bilger, and Deatherage (1956) have reported that estimates of durations up to 16 seconds were systematically distorted if the noise levels during the presentation of the interval and during the *S*s' attempt at reproducing that interval were different. The change was in the direction of overestimating the interval when the noise level was higher during the reproduction period and underestimating the interval when the reverse was the case.

Jerison and his co-workers (Jerison, Crannell, & Pownall, 1957) also found an effect of noise on time estimation. Two hundred *S*s, working individually, were required to follow a moving target visually and to imagine the continuing movement of the target after it disappeared. When the target was believed to have reached a crosshair, the *S* squeezed a trigger. Noise of about 110 db (frequency range: 20–10,000 cps) was introduced at certain times. It was found that a noise program in which it was quiet when the target disappeared gave longer judgment times relative to those obtained under control conditions of quiet or noise throughout. The opposite program of "noise then quiet" was not differentiated from the control conditions. It was also found that judgment times became longer in succeeding trials under all four noise conditions. An effect of noise on subjective time judgments has also been noted by Loeb (1957).

In addition to these recent studies that indicate an effect of noise on certain kinds of performance, there have been a number of observations on the effects of intermittent noise on performance. Many years ago Cassel and Dallenbach (1918) found that an intermittent noise resisted habituation more than a continuous one. Similarly, Laird (1933) reported that *S*s who were required to put a stylus through small irregularly spaced and sized holes in a moving tape, showed that the drop in output was greater when the noise distractor was made to vary in intensity in an eight-second cycle than when its loudest component was presented as a steady sound.

In two recent studies, K. R. Smith (1950, 1951) compared the performance of two groups of college students on number checking, name checking and form board tests. The experimental group was exposed to a series of 100 db intermittent sound bursts of 10 to 50 seconds in length, but so arranged that the total noise time was equal to the total silence time during the 30-minute period of the tests. It was found that the experimental group tried more items, scored more correctly, scored more incorrectly, and was less accurate than the control group. Some of the differences were small but significant. One other interesting fact emerged: with two exceptions, the experimental group was more homogeneous (i.e., smaller *SD*s) than the control group.

Another experiment which reports an effect of intermittent noise on performance is that of Corso (1952). He used a 100 db, 100–3000 cps noise, which was introduced intermittently throughout the test period while the *S*s worked on the Minnesota Clerical Test and the Minnesota Form Board Test. Here, just as in the previously cited experiment by K. R. Smith, the *S*s attempted more test items, got more correct, and made more errors, even though their performance was reported as less variable under the noise than under the control conditions. (Those who had performed well on the tests under control conditions tended to do poorer under the

stress, while on the average, those who had performed poorly in the control session improved their scores under stress.)

In summary then, it is possible to show that recent research indicates that high intensity noise has an effect on certain types of complex tasks, and that intermittent noise, in the few cases where it has been studied, seems to have a greater tendency to impair performance than steady noise.

On the basis of this past research, three hypotheses are offered:

1. Performance impairments are more likely to occur during high intensity intermittent sound than under lower intensity or steady sound.

2. The relative difficulty of the task is an important variable determining the effect of noise on performance. The more difficult the task, the more likely will noise be disruptive.

3. Individual differences in reaction to noise are extremely important to consider in evaluating research. There is evidence to indicate (Plutchik, 1955) that the mean of a group's responses may be misleading if there are marked individual differences or subgroups within the larger group. It is hoped that these hypotheses will be systematically tested in subsequent research.

One other major question is involved in a study of the effects of noise on performance. It is evident that a high intensity noise of whatever form can be considered a *stress* stimulus, and the literature dealing with stress should have relevance to noise and performance studies. A recent review and critique of studies of stress (Harris, Mackie, & Wilson, 1956) presents the following findings and conclusions:

1. There are wide individual differences in reactions to stress. The reasons for these different reactions have not been clearly identified.

2. The majority of the studies have been concerned with the effects of relatively short-term stress conditions, which means that temporary compensatory performance is possible. In many cases, the durations of the stress stimuli have not been reported.

3. Apparently no one has systematically investigated the relation between the intensity of the stress stimulus and its effect on behavior.

4. Few investigators have attempted to observe the course of behavior during stress.

5. Of the many different experimental designs which have been used, it is suggested that the subject-control design which compares a person's performance under both stress and non-stress conditions, is more satisfactory than a random group or group-control design method.

In the light of these observations it is evident that much further research is needed to clarify inconsistencies and to develop a theoretical rationale for the results which have been reported.

The Effects of Intermittent Sound on Feelings

There are at least three aspects to this problem that require examination: (*a*) the nature of the feelings that are associated with high intensity noise; (*b*) the effect of such noise on threshold, adaptation, and auditory acuity; and (*c*) the special subjective characteristics of intermittent, repetitive, or pulsed sounds.

With regard to the first aspect, the literature is almost unanimous: high intensity noise, even when it may have no effect on performance, will generally produce symptoms of discomfort, irritability, and distraction.

To illustrate: In the Benox report (Miles, 1953) a steady noise of 115 db heard for three-hour periods produced fatigue and discomfort; Miller (1951) had his *S*s listen to 111 db noise for 30-minute periods and they reported irritation, distraction, and general disturbance. Blau (1951) reports that a high intensity noise source of 103 db which accompanied the administration of various tests had practically no effect except to tend to arouse "somatic complaints of specific anatomical location and description," and Mendelson and Conway (1947) who exposed 10 volunteers to jet engine noise for 14 days and a total of 19 hours (at about 120 db over the range of 20–8000 cps), found that seven of the 10 *Ss* reported fatigue, irritability, and nervousness.

In spite of this agreement on the subjective effects of high intensity noise, there have been very few studies designed to isolate the more disturbing aspects of the noise spectrum. One of the few dealing with this problem studied the relative annoyance produced by various bands of noise (Reese & Kryter, 1944). The authors used filter systems to divide "white" noise into several bands, and five *S*s were asked to adjust each band of noise to equal a standard band (of 1900 to 2450 cps at 94 db) in "annoyance" value. It was found that frequencies above 2000 cps were more annoying than those below it. The results showed that "annoyance" as a characteristic of sound is discriminable from loudness, although with continued testing the annoyance and loudness contours became less separated.

One other early study dealing with this problem (Laird & Coye, 1929) had 14 *S*s compare the relative annoyance of eight different frequencies ranging from 64 to 8192 cps, by the method of paired comparisons. A U-shaped curve resulted with frequencies of 256 to 1024 least annoying, with a marked rise in relative annoyance for lower and higher frequencies. The authors conclude that at low intensities annoyance is approximately proportional to loudness while at high intensities (over about 80 db), the frequencies below 500 cps follow the equal loudness curves, while frequencies over 500 are equally annoying at lower loudness levels. Thus, annoyance is different from loudness per se.

Kryter summarized the findings dealing with the annoyance value of various noises by citing the aspects of a sound which tend to affect annoyance value. These are: (*a*) unexpectedness, (*b*) inappropriateness, (although this is a term which is difficult to specify), (*c*) intermittency (irregular, variable sounds are more annoying than steady ones), (*d*) reverberation (lack of localizing increases annoyance), (*e*) loudness (the threshold at which any sound becomes annoying has not yet been unequivocally determined), (*f*) frequency pattern (sounds having their energy concentrated in the higher audible frequencies are more annoying).

So far as it has been possible to determine, the relative annoyance value of different kinds of intermittency has not been studied.

Fatigue Effects of High Intensity Sound

In spite of many reports on the after effects of high intensity noise, including the extensive study by Davis (1942) during World War II using intensities up to 130 db and durations as long as 64 minutes, the problem has not yet been entirely

clarified. In general, it is known that in auditory fatigue, duration of stimulation has a cumulative effect from 30 seconds up to at least 10 minutes (Hallpike & Hood, 1941; Harris, 1953); that the maximum fatigue effects of a given frequency may be a half octave higher (Davis et al., 1943); and that there are very marked individual differences in susceptibility to auditory fatigue (Harris: 1953, 1954; Wilson, 1950). The existence of marked individual differences are attested by the fact that the least susceptible *S*s may return to normal in less than seven minutes from stimulation which the most susceptible do not recover from in more than 24 hours (Harris, 1953). Hearing loss, according to Harris (1953) tends to be a linear function of both stimulus duration and stimulus intensity, but for fatigue-resistant individuals recovery from 10 minutes of exposure to noise levels of 120 db takes only about seven minutes. Hearing loss for relatively short exposures, seems to be completely reversible. For example, a recent study (Thwing, 1956) reported that the adaptation produced by a 70 db tone of 1000 cps presented for six minutes is followed by complete recovery in about one minute after the termination of the adapting stimulus.

A test for screening fatigue-susceptible individuals has been reported in the literature by Wilson (1950). He exposes the *S* to a 2048 cps tone at 80 db intensity for five minutes, and threshold shifts great enough to prevent the *S* from hearing a tone 10 db over his own threshold after two minutes recovery are used as a criterion of fatigue. Harris (1954) suggests that a better criterion is either the time in seconds to return to within 5 db of own threshold, or the residual hearing loss after one minute. In both studies, the test tone is presented at 4096 cps since this is the region most affected by stimulation at 2048 cps. It is of some interest to note, in this connection, that when white noise, with a frequency band over most of the auditory spectrum, was used at 100–120 db for several minutes, the frequency of maximum fatigue, for all observers, was reported at between 4000 and 8000 cps (Hirsh & Ward, 1952).

Another more recent test of noise susceptibility has been proposed by Jerger and Carhart (1956) on the assumption that an ear's tendency to develop a permanent hearing loss is related to that ear's reaction to temporary acoustic stress. They found that out of 178 Air Force jet-mechanic trainees only 15 (or eight per cent) showed a hearing loss of 10 db or more on audimeter tests in the 3000–4000 cps range, eight weeks after a three day (12 hour) exposure period to jet-engine noise (SPL level unspecified, but probably in the range of 120–140 db). There was a slight but significant positive correlation between the amount of temporary threshold decrease to a 3000 cps, one minute duration tone, at 100 db, and the hearing loss found after eight weeks.

Several relatively recent studies have attempted to determine the effects of jet-engine noise on hearing loss. Mendelson and Conway (1947) used 10 *S*s exposed to sound intensities of about 120 db for a total of 19 hours, and reported that hearing losses of 20 to 60 db at frequencies of 512 to 2048 cps vanished after a week-end's rest and did not return. Davis et al. (1953) exposed 17 men to 21 bursts of jet-engine noise (at 126–150 db) each lasting 15 seconds, and separated by 30–45 second intervals

of relative silence. Audiograms taken two to three hours after exposure showed no temporary loss in any *S*, although the authors point out that people who come in constant contact with jets show definite hearing losses. Eldred and his co-workers (Eldred, Gannon, & Gierke, 1955) found that a one minute exposure to jet noise at 130–135 db produced a slight hearing loss followed by complete recovery within eight hours.

In the Benox report (Miles, 1953) a study is cited by Silverman (1947) in which he was able to locate the pain threshold for both normal and hard of hearing ears at about 140 db using earphones. The sensation of pain ("it hurts") was distinguished clearly from sensations of auditory discomfort ("it is too loud"), and of touch ("it tickles," or "I feel something in the ear").

The Benox researchers compared the pain threshold in a free field, using jet engines as noise source, with pain threshold determined by earphones and concluded that they were the same. They also noted that in the frequency range between 800 and 2000 cps the sound became "uncomfortably loud" at sound levels well below the pain threshold. This latter finding is consistent with the work of Hardy (1952) who points out that certain sounds, even 85 db, if prolonged over months and years may cause some degree of deafness. He concludes that sounds which exceed 100 sones per octave band are probably damaging with long-time daily exposure but that no damage is expected if no octave band exceeds 50 sones.

It may be concluded from these various observations and reports that the short time exposures to high intensity noise levels in the laboratory are not likely to produce any kind of permanent hearing losses, and that the effects which do appear are very transient, although there may be some marked individual differences in the speed of recovery.

Some Subjective Characteristics of Repetitive Sounds

When pure tones are presented repetitively for very brief durations, then at least two important phenomena come into existence. If the duration of each tone is extremely short, the tone is heard as a click which does not have a discernible frequency. As the duration is gradually increased a "click-pitch" threshold is found (i.e.) the shortest duration of a tone which allows the tone to have some pitch character to it), and then a "tone-pitch" threshold is found (the shortest duration of a tone in which the tonal character, rather than the click character is dominant). The exact values of these thresholds decrease with frequency, within the range of 125–8000 cps, but in no case is it greater than 18 milliseconds for the click-pitch threshold, or 25 milliseconds for tone-pitch (Doughty & Garner, 1947).

The second phenomenon referred to is a result of the abrupt presentation or removal of a tone. If a pure tone is interrupted at a given frequency, not only is the original tone obtained but also additional frequencies which are a function of the rate of interruption. A square wave pulse modulation produces a spectrum of sideband frequencies with maximum energy in the central component. Changing the repetitive rate or the duration of a pulse changes the spectrum of energy. According to Garner (1947b) energy changes at the rate of 6 db in the central component for every doubling or halving of either the repetition

rate or duration; each doubling or halving of the repetition rate or duration changes the total energy by 3 db. The threshold response is primarily determined by the duration of the pulse, and not the repetition rate, for low rates. This spectrum of sideband frequencies adds a *click* to any pure tone that may be used.

Although some studies (Garner: 1948, 1949) have made no attempt to eliminate or deal with the click introduced by the abrupt onset or decrease of a tone, two remedial procedures have been mentioned in the literature. One way to suppress the click due to switching the tone on and off is to produce a gradual increase and decrease of the tone, rather than to use a square wave (Luscher & Zwislocki, 1949); Munson (1947) used a three millisecond rise and fall time for this purpose while Miller and Heise (1950) used a 20 millisecond rise and fall time. Another way of dealing with this problem is to use a wide-band noise source with a uniform spectrum ("white-noise") as the interrupted tone because the sidebands produced by the interruptions fall within the existing spectrum and do not produce a change in the subjective character of the tones. This procedure was used by Pollack (1941) and by Miller and Taylor (1948).

Interestingly enough, the early study by Shower and Biddulph (1931) compared the effect on relative discrimination of the transients introduced by abrupt switching on or off of a tone. They found that the only effect was a decrease in the Weber ratio for frequencies below 500 cps. This indicates that the effect, if any, of a click cannot be assumed to necessarily disrupt or inhibit some auditory function, without investigation.

When tones are presented repetitively in brief pulses, there are several parameters that need to be specified in addition to the intensities and frequencies used. There is the *duration* of the tone, the *repetition* rate, the percentage of the total time that the tone is on, which represents an *on-off ratio*, and the total energy in decibels (Garner, 1948). Miller (1948) has called the on-off ratio by another name, the sound-time fraction, in a study of auditory "flutter" In his study he reports that in general the same types of relations for auditory "flutter" hold as do those for visual flicker with the exception that the critical flutter frequencies (135 bursts per second for one *S* and 270 bursts per second for the other) are much higher for fusion than are visual flicker rates.

In an extension of this work on auditory flutter, Pollack (1952) used flutter rates from 0.4 to 200 bursts per second, with five *S*s. When the relative change in flutter rate (that is, $\Delta f/f$) is plotted against flutter frequency, a minimum is found in the region of 10 per second. Pollack hypothesizes that this is not unrelated to the fact that the alpha rhythm of the brain is also about 10 per second.

Several interesting studies have been reported dealing with the ability to count short repetitive pulses, and with the loudness of pulses in comparison to steady tones. Taubman (1944), for example, had his *S*s judge the number of dots that would be sounded while different numbers of dots, from one to six were presented at either 10 per second, 14 per second or 18 per second rates. He found relatively little difference in the number of dots judged at these three different rates, although knowledge of results

seemed to help. His data suggest that these three rates cannot be easily discriminated.

Consistent with this finding is the work of Cheatham and White (1954) and Garner (1951). Cheatham and White used a 1000-cycle tone presented in pulses of 10 per second, 15 per second, and 30 per second, with each pulse of 11-millisecond duration and 70 db intensity. For four *S*s, it was found that regardless of the objective rate used, the subjective rate for auditory perception of sound pulses approaches a limit of about 9 to 11 pulses per second. In addition it was noted that the variability of response increases suddenly for mean perceived numbers higher than five. Garner used a similar task and procedure, and reports the following conclusions: (*a*) the duration (5–40 msec) and intensity (up to 94 db) of the tone had no effect on counting accuracy; (*b*) the curve for a repetition rate of 12 per second is almost identical with that for a 10 per second rate; and (*c*) there are very large individual differences in counting accuracy.

Garner has also reported in another study (1947b) dealing with threshold in relation to repetition rate, that as the repetition rate increases, the threshold decreases, for all frequencies used (250–4000 cps) although there is a break in the curve at rates between two and five tones per second. Garner notes that although the total energy in a stimulus is directly proportional to the repetition rate, it has been shown that the ear does very little integrating of acoustic energy beyond a duration of 200 milliseconds (or five pulses per second).

Two other studies are relevant to this problem. Mowbray and his coworkers (Mowbray, Gebhard, & Byham, 1956) used 10 interruption rates of white noise, ranging from 1 to 320 bursts per second at an intensity of 70–80 db. These pulse rates were given to the *S*s as standards and were to be matched by a variable pulse rate. When the average deviation of the matchings were plotted against pulse rate on logarithmic coordinates, two functions were revealed, suggesting a change in the method of frequency discrimination at about five bursts per second. This was interpreted as meaning that the listener is able to count the noise bursts from one to five per second but not above five.

This interpretation is nearly but not quite consistent with the results of an experiment by Licklider and others (Licklider, Webster, & Hedlun, 1950) dealing with the threshold for disappearance of binaural beats produced by pure tones. These experimenters found that the subjective character of beats changes in the range between 2 and 10 beats per second from periodic fluctuation to "roughness," depending upon frequency. When *S*s were asked to count the beats, the curve relating threshold to frequency was nearly a straight line at 8 beats per second. These studies, therefore, suggest that the counting ability of a human observer of tones, pulses, or beats has an upper limit definitely less than 10 and probably less than 8 per second.

In addition to these facts concerning the ability to discriminate pulse frequencies, there are several reports on the loudness of brief repetitive tones. Garner (1948) had six *S*s equate the loudness of a series of repeated short tones with the loudness of a steady tone of the same frequency using an 80 db intensity

level. His results showed the following:

1. A series of pulses may be louder than a steady tone of the same intensity level.

2. There is a maximum difference in loudness, in favor of the repeated short tones, at about 60 db. At higher and lower intensity levels, the relative advantage in loudness for pulse is less.

3. The loudness of pulses is relatively highest between 1000 and 4000 cps.

4. The most consistent matching of loudness is shown at 5–10 pulses per second, although a decrease in pulse duration tends to increase variability of judgments.

5. Loudness increases much less rapidly than does intensity.

In another related experiment dealing with binaural loudness matching with short tones, Garner (1947a) reported that the accuracy of loudness matching is best *both* below and above tone durations of 20 milliseconds (with repetition rates of 1–10 per second), and that repetition rates from 1–100 pulses/second had little effect on the variability of loudness matching. He notes as do many other investigators, that the differences between individuals are much greater than the differences within individuals.

In a study designed to study the loudness of noise rather than pure tones, Pollack (1941) reported findings somewhat similar to Garner's. He found that the interrupted noise sounds louder than a continuous noise of the same energy, being relatively greatest at rates of 2–10 pulses per second; and that the absolute threshold of a white noise (of constant sound-time fraction) increases as the frequency of interruptions increases from 1–300 pulses per second. Pollack points out that the range of 2–10 pulses per second is the range wherein there is enhancement of visual brightness of a flickering light; it is also where the most acute intensity discrimination is found between two tones as a function of the difference in frequency between the tones. This same range also covers the alpha rhythm and its first subharmonic.

In summary it may be concluded that humans have certain limitations with regard to the counting of rapid pulse signals. These limits include a range of about 1–12 pulses per second, a range in which certain interesting threshold, loudness, and variability data can be found. It may be stated as an hypothesis that such findings concerning auditory perception may be related to the alpha rhythm and various visual phenomena associated with it.

Physiological Effects of Loud Sounds

There are many published reports which indicate an effect of loud or repetitive sounds on physiological function. In his review of 1950, Kryter cites two earlier studies, one which found a rise in blood pressure following a loud unexpected sound (Lovell, 1941), and the other, which reported a decrease in peristaltic contractions and flow of gastric juices (Smith & Laird, 1930) following two 10-minute periods of noise.

More recently, Corso (1952) has stated that *S*s exposed to intermittent noise (800–10,000 cps) at 105 db while taking an intelligence test showed a significant increase in pulse rate during the noise condition, and Stambaugh (1950) reported the same result. Davis (Davis, Buchwald, & Frankman, 1955) found that when auditory stimuli near the pain threshold in intensity (about 120–130 db)

are presented to an *S* at one-minute intervals, a complex pattern of responses occurs, which includes increases in palmar-sweating, EMG, and respiration amplitude, decreases in pulse and finger volume and respiration rate, and an increase and then a decrease in pulse rate.

In addition, several studies have found a definite effect of loud sounds on muscle tensions. Davis (1948) used a 500 cps tone at an intensity level of about 90–100 db, presented to the *Ss* for two- or four-second intervals every two minutes. Although the *Ss* were instructed to do nothing in response to the stimuli, definite muscle tension responses were recorded which were partially dependent in magnitude on the level and geographic distribution of tension already existing in the *S* before the stimuli were applied. In a further development of this study, Davis (1950) found that *Ss* made slight muscular responses to subliminal auditory stimuli as well as to supraliminal sounds. In an extension of this type of experimental study to clinical patients, Malmo and his coworkers (Davis, Malmo, & Shagass, 1954; Malmo, Shagass, & Davis, 1950), (using 3-second, 1000 cycle tones at 80 db, 90 seconds apart) found that a group of psychoneurotics with extreme anxiety showed a greater and more long-lasting electromyographic response from the forearm than a control group of normal *Ss* and that arm muscles showed more significant differences than head muscles. There were, however, no significant differences between schizophrenic and control *Ss*.

It should be noted that in the studies by Davis there was some decrease in the muscle tension responses with continued testing, although this was not reported in Malmo's study.

In three investigations of the biological effects of jet-engine noise, two reported positive effects, and one reported none. Parrack (Parrack, Eldridge, & Koster, 1948) exposed men to 10-minute periods of jet-engine noise at 120–150 db and to a 1300 cps siren; the *Ss* reported heating of the skin, vibration in parts of the body, muscular weakness and excessive fatigue, all of which disappeared within a week after cessation of tests. Allen, Frings, and Rudnick (1948) using 10-second exposures of a 20,000 cps tone at 160–165 db (re: 10^{-16} watts per sq. cm.) reported that flies, mosquitoes, roaches and caterpillars were killed within minutes, while human *Ss* developed skin burns, slight dizziness, and unusual fatigue. In contrast to these two reports, Finkle and Poppen (1948) reported no measurable effect on a wide variety of physiological measures of one-hour periods of exposure for 10 days and two-hour periods for five days, to jet-engine noise at about 120 db. The difference in results between this study and the preceding two may be due to the lower intensity noise and to the difference in the distribution of energy in the different frequency ranges.

Several other reports have been made of the effect of sound on physiological processes. Mendelson (1957) measured the auditory reflex in man as a minute pressure change in one ear canal as a function of the intensity (80–115 db) and frequency (400–8000 cps) of tones in the other. Seven out of 23 *Ss* showed a definite effect. Coleman and Krauskopf (1956) reviewed the previous reports of an effect of loud sounds on vision and indicated that their experiments, using 110 and 140 db tones, revealed no consistent influences on visual threshold. They criticized the previous

findings as being based mainly on subjective impressions.

In the Benox Report, Ward (1953) presented the results of observations of the electroencephalograph of one *S* while he was exposed to siren frequencies of 245–700 cps at 120–137 db. When the earplugs were removed from one ear, there was a "striking and prompt abolition or desynchronization of the parieto-occipital alpha rhythm, at the same time good activation of the EEG in the frontotemporal region bilaterally was present, consisting of desychronization with an increase in the low potential, fact activity. When the eyes were open, the addition of the siren noise to one ear, added nothing to the alpha blockade already present. These EEG effects appeared to become less marked with repeated opening of the ear during any given run, suggesting the possibility that adaptation of central neural circuits can occur under these conditions." It is unfortunate that no other *S*s were tested since the two other studies which measured EEG's (Finkle & Poppen, 1948; Mendelson & Conway, 1947) during exposure to jet-engine sounds report no, or equivocal, changes in EEG recordings.

There is another important area of research to which attention should be drawn, although it might seem quite peripheral; this is the work dealing with sound-precipitated convulsions in animals. The most recent review of the literature by Bevan (1955) covering the period from 1947 to 1954, included 145 titles in the bibliography. The major findings of relevance here seem to be:

1. Pure-tone experiments would indicate that the most effective frequencies to produce convulsions in rats and mice, lie above 8000 cps, and the most effective intensities over 100 db.

2. At least one author has found that an intermittent pure tone (9500 cps) with a three-per-second interruption rate (a rate similar to that of certain biolectric potentials in epilepsy) was most effective in producing convulsions.

3. Stimulation of other sense modalities without sound does not seem to produce seizures.

4. Many different variables, both in the *S*s and in the environment, influence the incidence and magnitude of seizures.

In the light of some data to be presented in succeeding sections, it is possible that the phenomenon of audiogenic seizure is one which is not limited to rats and mice, but that humans too may show dispositions toward such a response under certain conditions.

Physiological Effects of Repetitive Sounds

Relatively few studies have dealt directly with this problem although there are some which have great theoretical value; these, largely by Lovett Doust and his co-workers will be described in some detail. One of their first papers appeared in the British journal, *Nature* (Lovett Doust, Hoenig, & Schneider, 1952). They reported that the use of a flickering light on 25 normal *S*s at frequencies between 3 and 32 flashes per second produced marked changes in oxiometrically determined arterial blood oxygen-saturation values. Flicker rates between 3 and 9 per second, and also between 12 and 17 per second produced a decrease of blood oxygen-saturation values, while this was normal at 9–11 flashes per second, and elevated at frequencies

of 18–22 per second. The comment was made in this paper that similar changes can be produced by replacing the flickering light by a relatively low-intensity auditory stimulus modulated to the same frequency.

In their next, more extensive paper (Lovett Doust & Schneider, 1952) the authors pointed out the rationale for such studies:

> Biological rhythms exist in a rich variety and almost bewildering profusion to attend and equilibrate the physiology of man. Such dynamic phasic activity appears not only to be intimately concerned with the phenomenology of life and biological processes in general, but is also to be found in purely chemical systems. Modalities of the periodicities associated with life can be divided into those external to the organism—including diurnal and climatic variation, sun-spot activity, etc., and into those inherent within the individual such as the respiratory and cardiac rhythms, the menstrual cycle, sleep and awakening. Only less well marked are certain psychological periodicities such as "cyclothymic" variations in mood and personality. In the course of the present century much painstaking research has attempted to link external with internal rhythmic activities, significant correlations being adduced between seasonal variation and, for example, the incidence of psychiatric disorder, immunity from disease, temperament and behavior, and an impressive array of biochemical and physiological variables ranging from blood pH, lactic acid and protein to breath-holding time, plethysmography, tests of hand strength and fatigability, dark adaptation time and various tests of urinary function.
>
> Neurophysiologically, there is evidence both for chemical as well as for electrical patterns of periodic activity in the nervous system. Repetitive discharges with frequencies of 5–10 cycles per second from isolated stellate ganglion preparations were found to be capable of change by variation in the ionic concentrations of the surrounding medium and Kaufman and Hoagland have findings suggesting a chemical pacemaker closely identified with cerebral respiration. The function of the diencephalon as a neural pacemaker has long been postulated, and it appears certain that the hypothalamic aggregation of nuclei with its neural and endocrine connections must play an outstanding role in emotional and awareness variations (p. 640).

Using 109 *S*s, 58 normal controls and 51 hospital patients, they reported data showing that the arterial oxygen saturation levels vary very consistently (as described above) for normal *S*s regardless of the type of rhythmic stimulation used. (Photic, auditory, and cutaneous stimuli were used.) Maximum anoxemia (oxygen saturation decrements) occurred at 5 and 15 pulses per second in the normal *S*s and they showed in addition a summative effect of simultaneous sonic and photic stimulation. The patient group showed similar but not identical stimulation profiles although some showed unpredictable variations in response to the different frequencies.

"By depressing the oxygen levels by the choice of optimal stimulation frequencies, spontaneous comments by the healthy subjects revealed considerable changes in affect and levels of awareness, while, among the patients, repressed unconscious material was brought into consciousness." Some of the spontaneous comments made during the anoxemia periods included: "concentration poor, feel slowed up; tired, drowsy, sleepy; irritable, annoyed, fed up; desire to stop machine or break it; headaches, dizziness, or giddiness."

The third paper of the series (Lovett Doust, 1953) discussed the applications of these findings to an analysis of mental illness. Lovett Doust presented the hypothesis that an important feature of mental disorder, at least in the physiological sphere of reference is a relative anoxemia, either in the resting state or accompanying the dynamic response of psychiatric patients to stress situations. The fourth paper of the series

(Lovett Doust & Schneider, 1954) presented some preliminary data on a new treatment for patients with psychiatric disorders based on rhythmic sensory stimulation, a method which Doust calls "Rhythmic Sensory Bombardment Therapy" or R.S.B.T.

Regardless of the actual effectiveness of rhythmic sensory stimulation for the treatment of the mentally ill, this work is very important in showing a basic similarity of physiological effect of repetitive stimuli in different sense modalities. From this point of view, it makes sense to examine the studies of visual flicker as a (possibly) reliable guide to the effects of auditory flicker.

A visual phenomenon similar, in certain respects, to the one described by Lovett Doust, was discovered by Bartley in 1938. This is the fact of brightness enhancement, which is an increase in apparent brightness of a flickering light, at certain rates somewhat below those that bring about fusion. This has been studied by using two targets side by side—one steadily illuminated and the other intermittently. The intensity and rate of intermittency of the intermittent target are varied until the brightness of the two targets is matched. As the rate is reduced from the fusion point less and less intensity is required for the intermittent target to match the steady one. This continues until the pulse rate is in the neighborhood of 10 per second, the alpha rhythm of the human *S*. Here the intermittent target becomes about twice as effective as the steady one. With rates lower than this, the effectiveness of the intermittent target declines again. It has also been found that a light-dark ratio of one-to-one produces maximum relative enhancement and that the phenomenon occurs only with high intensity stimuli.

In addition to this Bartley effect, a flickering light produces many other results in human *S*s. At an ancedotal level, many people have unpleasant effects of flicker produced by driving through a forest or thicket of regularly planted trees which consistently and sequentially interrupt natural light sources; similar effects have been observed when the *S* stands still and observes a light through a rapidly rotating object such as a fan. In recent years, flickering lights have been used as an aid to diagnosis of "petit mal" epilepsy through observation of changes in electroencephalographic pattern. Most recent of all is the interest in "photic driving," as a possible clue to the nature of (hypothetical) reverberating circuits within the brain which are thought to be concerned with consciousness and perception.

In a fairly extensive and recent study of the effects of flickering light on human *S*s, Bach, Sperry, and Ray (1954) used high intensity light flashes at varying frequencies to determine the effects on subjective discomfort, tapping rates, walking, pursuit rotor performance, and rifle firing. Their results may be summarized as follows:

1. Unpleasant subjective effects are consistently reported when *S*s are exposed to diffuse flickering light. These effects can be grouped into four main categories:

(A) Interference with consciousness—e.g.: blank mind, drowsy, dizzy, paralyzed

(B) Sensations involving the eyes —e.g.: fatigue, irritated, watering

(C) Sensations of muscle twitching —e.g.: eyes blinking, facial twitching, jumping of body

(D) Sensations relating to other parts of body—e.g.: queasy

feeling in stomach, headache, nausea, chills, tense muscles. (Although the subjective reports in this experiment seem to have an unpleasant character, there is at least one other study in which pleasant sensations were also reported, for example, sensations of warmth which may become quite pleasurable, or illusory conversations, the content of which could not be recalled (O'Flanagan, Timothy, & Gibson, 1948).

2. The most consistently effective flicker-frequency for the production of subjective effects is 9 flashes per second, although any particular frequency is not extremely critical between the limits of 7 and 20 flashes per second. The effects are not cumulative with time of exposure beyond about five minutes, while maximum effects occur with high brightness of the field of view. Monochromatic light seems to hold no advantages over white light as indicated by preliminary tests with red, blue, and white light. Some degree of drowsiness was reported in *all* cases where the light was modulated by the spontaneous 9 cps EEG activity of the *S*.

3. Hand-eye coordination was significantly impaired by a flickering light for one task (tapping) but not affected on another supposedly more complex tapping task. Rifle firing accuracy was significantly depressed (by approximately 50%) when a 6 cps flickering light was placed behind the target at brightness levels of the order of magnitude of the scattered light of searchlight beams, but when the direction of the diffuse flickering light was toward the target, rifle firing accuracy increased. Rate of walking under conditions of diffuse flickering light did not seem to be significantly affected, even in the presence of obstacles and with a continuously moving light source.

This report therefore, as well as a host of previous studies (Ellingson, 1956) indicates clearly the marked effects which high intensity intermittent stimuli may have on behavior, feelings, and on certain physiological measures. In fact it has been recently possible to develop an index of anxiety proneness by examination of the EEG reactions to photic stimulation. Ulett et al. (1953) tested 191 *S*s, both patients and controls, on various psychological tests and by interviews, and had them rated for anxiety proneness under stress. EEG records were then taken during conditions of resting and intermittent photic stimulation. The results indicated that there was a significant correlation between the psychological criteria of anxiety proneness and: the percentage of fast, slow and low amplitude alpha in electronically analyzed resting EEG responses; the amount of harmonic responses in the 20–30 cps range when flicker frequencies of one half or one fourth of this were used; and the amount of subjective dysphoria produced by photic stimuli. A check list of EEG anxiety indicators derived from this correlated +0.51 with the validating criteria of anxiety proneness.

As far as could be determined there is only one study which has reported an attempt to "drive" brain rhythms by means of intermittent auditory stimulation. This study was described very briefly in an abstract of a talk to an EEG Congress by Goldman (1952). He reported that pure tones, which were interrupted in rhythmic fashion at rates of 1.5 to 25 per second, were used and that EEG changes appeared showing acoustic driving in temporal areas in

two out of eight cases. No other data were given and it is evident that repetition of this study would be interesting and fruitful.

Summary and Conclusions

This review has attempted to summarize and integrate a number of articles dealing with the effects of loud and intermittent sounds on human behavior, feeling, and physiology. Most of these studies have been published since 1950 when the last comprehensive review was written.

Some of the more recent experiments demonstrate effects of very loud sounds on certain kinds of complex behavior particularly "clock-watching" and time estimation with the possibility implied that the decrement in performance may depend on the level of the sound as well as on its intermittency.

With regard to the effects of high intensity or intermittent sound on feeling, the literature is almost unanimous: high intensity noise, even when it may have no effect on performance will generally produce symptoms of discomfort, irritability, and distraction, although there is little known about the relative annoyance value of different kinds and levels of intermittency.

Certain unique subjective characteristics of repetitive sounds are described relating to the effects of various rates of repetition on fusion, estimation of pulse frequency, tonal character, threshold, and loudness. The greatest effects are usually obtained at repetition rates between 5 and 10 pulses per second, a frequency range which coincides more or less closely with the alpha rhythm of the brain.

With regard to the effects of loud or intermittent sounds on physiology, changes in blood pressure, gastric secretion, pulse rate, palmar sweating, respiration, muscle tension, the electro-encephalogram, and blood-oxygen saturation have been reported in various studies.

Some theoretical concepts are presented which postulate effects of auditory intermittency parallel to those of visual flicker. In general, the need for more research on the effects of intermittent sounds of various frequencies, repetition rates, and intensities is evident from this review.

REFERENCES

Allen, C. H., Frings, H., & Rudnick, I. Some biological effects of intense, high frequency airborne sound. *J. acoust. Soc. Amer.*, 1948, **20**, 62–65.

Bach, L. M. N., Sperry, C. J. Jr., & Ray, J. T. Studies on the effects of flickering light on human subjects. New Orleans, La.: Tulane Univer., 1956. (Contract DA-44-009-ENG-2448)

Bartley, S. H. Visual response to intermittent stimulation. *Optical J. Rev.*, 1952, **89**, 31–33.

Berrien, F. K. The effects of noise. *Psychol. Bull.*, 1946, **43**, 141–161.

Bevan, W. Sound-precipitated convulsions: 1947 to 1954. *Psychol. Bull.*, 1955, **52**, 473–504.

Blau, T. H. Effects of high intensity sound on certain psychological variables. Unpublished doctoral dissertation, Penn. State Coll., 1951.(a)

Broadbent, D. E. The twenty dials and twenty lights test under noise conditions. A.P.U. 160/51. London Med. Res. Council, 1951.(a)

Broadbent, D. E. Noise, paced performance, and vigilance tasks. Med. Res. Council, 1951. (A.P.V. Rep. 165/51.)(b)

Broadbent, D. E. Noise, paced performance and vigilance tasks. *Brit. J. Psychol.* (Gen. Sec.), 1953, **44**, 295–303.

Broadbent, D. E. Some effects of noise on visual performance. *Quart. J. exp. Psychol.*, 1954, **6**, 1–5.

Cassel, E. E., & K. M. Dallenbach. The effect of auditory distraction upon the sensory reaction. *Amer. J. Psychol.*, 1918, **29**, 129–143.

CHEATHAM, P. G., & WHITE, C. T. Temporal numerosity: III. Auditory perception of number. *J. exp. Psychol.*, 1954, **47**, 425–428.

COLEMAN, P. D. & KRAUSKOPF, J. The influence of high intensity noise on visual threshold. Fort Knox, Ky.: Army Med. Res. Lab., 1956. (Rep. No. 222.)

CORSO, J. F. The effects of noise on human behavior. *WADC Tech. Rep.* 1952, No. 53-81.

DAVIS, H., ET AL. Physiological effects of exposure to certain sounds; final report. *OSRD Rep.* No. 889, 1942.

DAVIS, H., ELDREDGE, D. H., & GLORIG, A. Bioacoustic studies of high intensity aircraft engine noise at the Naval air-test center and aboard USS Coral Sea (CVA. 43). *Rep. on Biological Effects of Noise—Exploratory.* 1953. (ASTIA-DSC AD-47051).

DAVIS, H., MORGAN, C. T., HAWKINS, J. E., JR., GALAMBOS, R., & SMITH, F. W. Temporary deafness following exposure to loud tones and noise. Comm. Med. Sci., *OSRD*, 1943. (Rep. No. PBM 57250.)

DAVIS, J. F., MALMO, R. B., & SHAGASS, C. Electromyographic reaction to strong auditory stimulation in psychiatric patients. *Canad. J. Psychol.*, 1954, **8**, 177–186.

DAVIS, R. C. Motor effects of strong auditory stimuli. *J. exp. Psychol.*, 1948, **38**, 257–275.

DAVIS, R. C. Motor responses to auditory stimuli above and below threshold. *J. exp. Psychol.*, 1950, **40**, 107–120.

DAVIS, R. C., BUCHWALD, A. M. & FRANKMAN, R. W. Autonomic and muscular responses and their relation to simple stimuli. *Psychol. Monogr.*, 1955, **69**, No. 20, (Whole No. 405).

DOUGHTY, J. M. & GARNER, W. R. Pitch characteristics of short tones. I. Two kinds of pitch threshold. *J. exp. Psychol.*, 1947, **37**, 351–365.

ELDRED, K. M., GANNON, W. J., VON GIERKE, H. E. Criteria for short-time exposure of personnel to high intensity jet aircraft noise. *WADC Tech. Note* 55-355. Ohio: Wright-Patterson Air Force Base, 1955.

ELLINGSON, R. J. Brain waves and problems of psychology. *Psychol. Bull.*, 1956, **53**, 1–34.

FINKLE, A. L. & POPPEN, J. R. Clinical effects of noise and mechanical vibrations of a turbo-jet engine on man. *J. appl. Physiol.*, 1948, **1**, 183–204.

GARNER, W. R. Accuracy of binaural loudness matching with repeated short tones. *J. exp. Psychol.*, 1947, **37**, 337–350.(a)

GARNER, W. R. Auditory thresholds of short tones as a function of repetition rates. *J. acoust. Soc. Amer.*, 1947, **19**, 600–608.(b)

GARNER, W. R. The loudness of repeated short tones. *J. acoust. Soc. Amer.*, 1948, **20**, 513–527.

GARNER, W. R. The loudness and loudness matching of short tones. *J. acoust. Soc. Amer.*, 1949, **21**, 398–403.

GARNER, W. R. The accuracy of counting repeated short tones. *J. exp. Psychol.*, 1951, **41**, 310–316.

GOLDMAN, D. The effect of rhythmic auditory stimulation on the human electroencephalogram. *EEG. clin. Neurophysiol.*, 1952, **4**, 370. (Abstract)

HALLPIKE, C. S., & HOOD, J. D. Some recent work on auditory adaptation and its relation to the loudness recruitment phenomenon. *J. acoust. Soc. Amer.*, 1941, **23**, 270–274.

HARDY, H. C. Tentative estimate of a hearing damage risk criterion for steady state noise. *J. acoust. Soc. Amer.*, 1952, **24**, 756–761.

HARRIS, J. D. Recovery curves and equinoxious exposures in reversible auditory fatigue following stimulation up to 140 db plus. *Laryngoscope*, 1953, **63**, 660–673.

HARRIS, J. D. The roles of sensation level and of sound pressure in producing reversible auditory fatigue. *Laryngoscope*, 1954, **64**, 89–97.

HARRIS, W., MACKIE, R. R., & WILSON, C. L. Performance under stress: A review and critique of recent studies. *ONR, Personnel and Training Branch. Tech. Rep.* 6, 1956.

HIRSH, I. J., BILGER, R. C., & DEATHERAGE, B. H. The effect of auditory and visual background on apparent duration. *Amer. J. Psychol.*, 1956, **69**, 561–574.

HIRSH, I. J., & WARD, W. D. Recovery of the auditory threshold after strong acoustic stimulation. *J. acoust. Soc. Amer.*, 1952, **24**, 131–141.

JERGER, J. F., & CARHART, R. Temporary threshold shift as an index of noise susceptibility. *J. acoust. Soc. Amer.*, 1956, **28**, 611–613.

JERISON, H. J. Differential effects of noise and fatigue on a complex counting task. *WADC Tech. rep.* 1956, No. 55-359.

JERISON, H. J., CRANNELL, C. W., & POWNALL, DOROTHY. Acoustic noise and repeated time judgments in a visual movement projection task. *WADC Tech. Rep.* 1957, No. 57-54.

JERISON, H. J., & WING, S. Effects of noise and fatigue on a complex vigilance task. *WADC Tech. Rep.*, 1957, No. 57-14. Yellow Springs, Ohio: Aero Medical Laboratory, Wright Air Development Center,

KRYTER, K. D. The effects of noise on man. *J. Speech Dis. Monogr.* Suppl. No. 1. 1950.

LAIRD, D. A. The influence of noise on production and fatigue as related to pitch, sensation level and steadiness of noise. *J. appl. Psychol.*, 1933, **17**, 320–330.

LAIRD, D. A., & COYE, K. Psychological measurements of annoyance as related to pitch and loudness. *J. acoust. Soc. Amer.*, 1929, **1**, 158–163.

LICKLIDER, J. C. R., WEBSTER, J. C., & HEDLUN, J. M. On the frequency limits of binaural beats. *J. acoust. Soc. Amer.*, 1950, **22**, 468–473.

LOEB, M., & RICHARD, R. G. The effects of intense stimulation on the perception of time. Fort Knox, Ky.: *AMRL Rep.* 269, 1957.

LOVELL, G. D. Physiological and motor responses to a regularly recurring sound. *Psychol. Bull.*, 1941, **38**, 715. (Abstract)

LOVETT DOUST, J. W. Physiological factors underlying therapeutic success in mental disorders. *Canad. Med. Assoc. J.*, 1953, **69**, 108–114.

LOVETT DOUST, J. W., HOENIG, J., & SCHNEIDER, R. A. Effect of critical flicker frequencies on oximetrically determined arterial blood oxygen-saturation levels. *Nature*, 1952, **169**, 843–845.

LOVETT DOUST, J. W., & SCHNEIDER, R. A. Studies on the physiology of awareness; the effect of rhythmic sensory bombardment on emotions, blood oxygen-saturation and the levels of consciousness. *J. Ment. Sci.*, 1952, **48**, 640–653.

LOVETT DOUST, J. W., & SCHNEIDER, R. A. Rhythmic sensory bombardment therapy (R.S.B.T.), *Dis. Nerv. Sys.*, 1954, **15**, 3–15.

LUSCHER, E., & ZWISLOCKI, J. Adaptation of the ear to sound stimuli. *J. acoust. Soc. Amer.*, 1949, **21**, 135–139.

MALMO, R. B., SHAGASS, C., & DAVIS, J. F. A method for the investigation of somatic response mechanisms in psychoneurosis. *Science*, 1950, **112**, 325–328.

MENDELSON, E. S. Effects of jet-engine afterburner noise: Objective measurement of the auditory reflex in man. Phila., Pa.: Naval Air Material Center, 1957. (NAMC-ACEL-327.)

MENDELSON, E. S., & CONWAY, H. Report on Turbo-Jet Engine Noise. Phila., Pa.: Aero Medical Equipment Laboratory. U. S. Naval Base Station, 1947.

MILES, W. R. Immediate psychological effects. In Benox Report: An exploratory study of the biological effect of noise. Univer. Chicago, 1953. (*ONR. Project* NR 144079.)

MILLER, G. A., & HEISE, G. A. The trill threshold. *J. acoust. Soc. Amer.*, 1950, **22**, 637–638.

MILLER, G. A., & TAYLOR, W. G. The perception of repeated bursts of noise. *J. acoust. Soc. Amer.*, 1948, **20**, 171–182.

MILLER, H. G. A study of the interference effects of high intensity sound on the retention of visually and auditorilly learned material. Unpublished doctoral dissertation. Penn. State Coll., 1951.

MILLER, J. G. The development of experimental stress-sensitive tests for predicting performance in military tasks. Wash., D. C.: Psych. Res. Assn., 1953. (*Tech. Rept.* No. PBR 1079, Army Proj. No. 29562000).

MOWBRAY, G. H., GEBHARD, J. W., & BYHAM, C. L. Sensitivity to changes in the interruption rate of white noise. *J. acoust. Soc. Amer.*, 1956, **28**, 106–110.

MUNSON, W. A. The growth of auditory sensation. *J. acoust. Soc. Amer.*, 1947, **19**, 584–591.

O'FLANAGAN, P. M., TIMOTHY, J. I., & GIBSON, H. G. Further observations on the use of combined photic and chemically induced cortical dysrhythmia in psychiatry, *J. Ment. Sci.*, 1951, **97**, 174–190.

PARRACK, H. O., ELDREDGE, D. H., & KOSTER, H. F. Physiological effects of intense sound. USAF Air Materiel Command, Wright-Patterson Air Force Base, 1948. (*Aero Med. Lab. Rep.* No. MCREXD-695-71B.)

PLUTCHIK, R. Some problems for a theory of emotion. *Psychosom. Med.*, 1955, **17**, 306–310.

POLLACK, I. On the threshold and loudness of repeated bursts of noise. *J. acoust. Soc. Amer.*, 1941, **23**, 646–650.

POLLACK, I. Auditory flutter. *Amer. J. Psychol.*, 1952, **65**, 544–554.

REESE, T. W., & KRYTER, D. D. The relative annoyance produced by various bands of noise. Harvard Psychoacoustic Laboratory, 1944. (Rep. 1C-65.)

SHOWER, E. G., & BIDDULPH, R. Differential pitch sensitivity of the ear. *J. acoust. Soc. Amer.*, 1931, **3**, 275–287.

SILVERMAN, S. R. Tolerance for pure tones and speech in normal and defective hearing. *Ann. Otol. Rhinol. Laryngol.*, 1947, **56**, 658–677.

SMITH, E. L., & LAIRD, D. A. The loudness of auditory stimuli which affect stomach contractions in healthy human beings. *J. acoust. Soc. Amer.*, 1930, **2**, 94–98.

SMITH, K. R. Intermittent loud noise and mental performance. Ohio: Wright-Patter-

son Air Force Base, 1950. (*AF Tech. Rep.* No. 6368.)

SMITH, K. R. Intermittent loud noise and mental performance. *Science*, 1951, **114**, 132–133.

STAMBAUGH, C. J. An investigation of certain individual differences under the stress of high intensity sound. Unpublished doctoral dissertation. Penn. State Coll., 1950.

TAUBMAN, R. E. The effect of practice with and without reinforcement on the judgment of auditory number. *J. exp. Psychol.*, 1944, **34**, 143–151.

THWING, E. J. Masked threshold and its relation to the duration of the masked stimulus. *J. acoust. Soc. Amer.*, 1956, **28**, 606–610.

ULETT, G. A., GLESER, G., WINOKUR, G., & LAWLER, A. The EEG and reaction to photic stimulation as an index of anxiety proneness. *EEG. Clin. Neurophysiol.*, 1953, **5**, 23–32.

WARD, A. A., JR. Central nervous system effects. Benox Report. An exploratory study of the biological effect of noise. Univer. Chicago, 1953. (*ONR Project* NR. 44079.)

WILSON, W. H. Determination of susceptibility to abnormal auditory fatigue. *Ann. Otol. Rhinol. and Laryngol.*, 1950, **59**, 399–405.

Received July 30, 1958.

Psychological Bulletin
Vol. 56, No. 2, 1959

METHODOLOGICAL CONSIDERATIONS IN THE CONSTRUCT VALIDATION OF DRIVE-ORIENTED SCALES

DONALD H. KAUSLER and E. PHILIP TRAPP
University of Arkansas

Studies concerned with the dynamogenic effects of anxiety, as measured by the Taylor Manifest Anxiety Scale (MAS), on learning and performance (e.g., Spence & Farber, 1953; Spence, Farber, & McFann, 1956; Taylor & Chapman, 1955; Taylor & Spence, 1952) have typically regarded the MAS as a discriminating measure of generalized drive (*D*). Evidence for the validity of the test is considered present whenever extreme groups, assigned on the basis of test score, perform in the direction hypothesized by related *D* theory. The particular theory involved has been treated in detail by Farber (1955), Taylor (1956), and Spence (1958).

Construct Validity and the MAS

Experimental validation of a test within a theoretical framework represents an attempt at construct validation. Basic to construct validation, as conceived by Cronbach and Meehl (1955), is the existence of a nomological network that relates the construct to observables and to other constructs. Extending the validity of a construct would involve, according to Cronbach and Meehl (1955), ". . . elaborating the nomological network in which it occurs, or of increasing the definiteness of the components" (p. 290). They further stated that the validation of a test claiming to measure a construct would require the existence of a nomological net surrounding that construct. The recent article by Spence (1958) attempted, in fact, to describe the nomological net surrounding the construct of emotionally based *D*.

In their criticism of the construct validity of the MAS, Jessor and Hammond (1957) questioned the extensiveness to which various aspects of the nomological net surrounding *D* have been investigated experimentally and alternative inferences disconfirmed. They accurately noted that MAS studies have concentrated specifically on the energizing component of *D* and have generally ignored other, perhaps equally important, aspects of the net. The Jessor and Hammond criticism is certainly germane not only to the MAS but to other scales purporting to measure *D*.

The present writers, moreover, are inclined to the critical view that even those studies dealing with the dynamogenic aspects of the net encompassing *D* have, in general, contributed little to a systematic construct validation of the MAS and other *D*-oriented scales. The literature in the area fairly abounds with studies yielding conflicting or equivocal results. Experiments on serial and paired-associates learning, in particular, have been conspicuously contradictory in their findings. On one hand, Spence, Farber, and McFann (1956), Spence, Taylor and Ketchel (1956), Taylor and Chapman (1955), Taylor and Spence (1955), and others (Montague, 1953; Ramond, 1953) have reported evidence consonant with the nomologies present in the *D* net; on the

other hand, Saltz and Hoehn (1957), Katchmar, Ross, and Andrews (1958), Sarason (1957a, and 1957b). Hughes, Sprague, and Bendig (1954), and others have provided evidence against the expectations of the theory. The present writers contend that a major difficulty hampering and confusing research in this area has been the tendency to confound interactional effects within the nomological net with certain methodological problems inherent in the research design. The express purpose of this paper is to point out some of these pertinent methodological problems and to suggest some possible ways of resolving them.

Methodological Problems in the Nature of the Task

According to the proponents of *D* theory, both *D* level and the relative strengths of competing responses must be accounted for in predicting the effect of *D* upon performance. Taylor (1956), for example, explicitly stated the position as follows: "In situations in which a number of competing response tendencies are evoked, only one of which is correct, the relative performance of high and low drive groups will depend upon the number and comparative strengths of the various response tendencies" (p. 304). In simple learning situations such as conditioning where a single response tendency is to be acquired, the prediction from *D* theory is straightforward. High *D* groups are predicted to condition at a faster rate than low *D* groups. Considerable experimental evidence would seem to support this prediction based upon the MAS. In verbal learning, and especially paired-associates learning, the prediction would be based upon the amount of interference or competition within the list. *D* theory would predict that with little intratask interference a high *D* group should be superior to a low *D* group; however, on tasks involving considerable intratask competition, the low *D* group would be predicted to be superior. The validity of these nomologies within the *D* net, incidentally, has been questioned by Hill (1957) on purely theoretical grounds.

Because of the *D* level-response competition interaction that has been postulated to operate within the nomological net, research relating *D* level as measured by *D*-oriented scales (chiefly the MAS) to performance on paired-associates tasks requires a "control" over the degree of competition within the task. The empirical results of such studies would then represent evidence for a construct validation of the scale. A crucial condition, therefore, in these validity studies is the degree to which *E* has controlled intratask interference. The typical procedure is to select nonsense syllables or adjectives in such a way that the similarity of the material and the association value of the material within the list can be conveniently manipulated. Similarity is usually manipulated by varying the letter content of items within the list, and association value is regulated by selecting syllables from Glaze's (1928) or Hull's (1933) tables with previously calibrated association values. By such manipulations, *E* derives a list which he considers representative of either a competitional or noncompetitional task. As Saltz and Hoehn (1957) point out, this procedure confounds amount of intratask competition with the difficulty of the task. They contend that an increase in response competition is accompanied by an increase in difficulty of the list. They conducted sev-

eral studies in which they attempted to partial out difficulty from competition. As a result, their findings failed to support *D* level-response competition theory.

In manipulating lists of nonsense syllables or adjectives in terms of similarity and association value for purposes of varying intratask interference and/or controlling for task difficulty, *E* seems to be operating under the basic premise that competition within a list is independent of *D* as measured by his *D*-oriented scale. We seriously question this premise particularly as it applies to the association value of nonsense syllables and adjectives. The calibrations of Glaze and Hull, it may be recalled, were based upon groups undifferentiated on any *D* dimension. It does not seem unreasonable to suspect that performance on *D* scales and number of associations to nonsense syllables may have some covariance. If true, lists constructed from such calibrations would not be comparable lists in terms of competition and difficulty for high and low *D* groups selected on the basis of scale score. Some adequate empirical demonstrations of independence between these variables are clearly needed.

Methodological Problems in Drive Measurement

Another important methodological problem encountered in construct validation studies on *D*-oriented scales pertains to the conditional definition of *D* which the use of such scales involves. These difficulties have been discussed by Jessor and Hammond (1957). They concluded with the statement: "When a construct implies a relationship between variables, these variables must be designated independently of any test of that relationship" (p. 169). The methodology commonly employed in studies involving *D* scales and complex learning tasks has departed grossly from this important qualification. The scale has been employed both to establish the validity of the construct (*D*) and simultaneously to establish the construct validity of the scale. Under such confounding dual purposes, failure of the data to fulfill the predicted outcome cannot be taken as substantive evidence for either an absence of construct validity in the scale or for an incorrect nomological net. The results of such studies serve mainly to confuse and cloud the issue.

As a way out of this protracted dilemma, the writers suggest that research in this area begin to utilize experimentally induced *D* states as controls for evaluating the effects of response-inferred *D* states. We further recommend positive, nonemotional, approach drives in preference to negative, emotional, avoidance drives, such as shock or threat of shock, which are fraught with so many unsettled theoretical problems in their own right. The nomological net surrounding the positive drives is more clearly defined and has received wider empirical support than is generally the case for the negatively based drives.

Since experimentally induced *D* would be considered as contributing to generalized *D*, predictions based on this form of *D* should be comparable to those based on *D* inferred from performance on a scale. For example, groups performing under high and low incentive conditions should be at least partially equated with high and low *D* groups selected on the basis of MAS scores. Predictions of this nature would extend beyond mere rate of acquisition in complex

learning and would include such diverse phenomena as intentional versus incidental learning, positive and negative transfer, extinction, etc. The kind of controls advocated here would not, of course, eliminate from consideration other interpretations pertaining to the construct validity of *D* scales. Predictions concerning emotionally based *D* upon an interfering response theory, as exemplified by Sarason and Mandler (Mandler & Sarason, 1952; Sarason, Mandler, & Craighill, 1952), or upon a habit theory, as exemplified by Hilgard (1953) and Child (1954), would be difficult to test discriminately by any method. The procedure propounded here has as its basic merit the elements for establishing a more well-defined baseline in an area besieged with many intricacies.

Two alternatives are suggested here for providing the type of control discussed above. The first consists of a replication of a carefully designed, well-conceived experiment in which an experimentally induced positive *D* had been manipulated and clear-cut results obtained. The replication would consist of repeating the task and procedure as closely as possible with groups selected as high and low scorers, respectively, on the *D* scale to be validated. Such replication would require a preliminary demonstration that competition within the material to be learned showed no covariance with score on the scale. A study in preparation by the writers exemplifies this approach. It consisted of a replication of a study by Bahrick (1954) in terms of material learned and procedure followed, but was extended to a different *D* dimension. Bahrick had demonstrated that a high incentive (financial reward) group displayed significantly greater intentional learning but significantly less incidental learning than a low incentive (no financial reward) group. The learning task consisted of the serial learning of geometric forms. Since the nature of the task is one that probably involves little intratask competition, it should differentiate high and low *D* groups selected by MAS score in a similar manner. That is, *D* theory would predict that high scorers on the MAS should show greater intentional learning but less incidental learning than low scorers.

We found, as in the Bahrick study, that the high *D* group (now identified by MAS score), comparable in intelligence and sex distribution to the low *D* group, displayed a significantly higher rate of intentional learning. Unlike the Bahrick study, however, the low *D* group did not display superior incidental learning. The results would therefore seem to conflict with the finding by Silverman and Blitz (1956) that high scorers on the MAS showed significantly less incidental learning on a serial list of nonsense syllables with low association values than low scorers. Silverman and Blitz interpreted their findings in terms of interfering responses correlating with anxiety. Unfortunately, the lack of information concerning the comparability of the list of syllables for the two groups and the lack of control groups performing on the same task under more operationally defined *D* conditions make it difficult to evaluate such results as either supporting or rejecting the construct validity of the MAS.

The second suggested alternative follows in a natural vein from the first. Studies that are directed at the construct validity of *D* scales should include within their research design, whenever possible, comparable groups performing the same task

as the experimental groups but under varying levels of an experimentally induced approach *D*. The performance of such control groups would then provide a more adequate basis for evaluating the effects of *D* inferred from scores on a *D* scale. A recent study by Katchmar, Ross, and Andrews (1958) illustrates this point nicely. They compared rate of learning for a coding task between groups differentiated in *D* level in terms of MAS anxiety, ego involvement, and failure-induced stress. Their design is certainly excellent as far as making relative comparisons between these three variations of emotionally based *D*. In their discussion, however, they state the results for high and low MAS scorers do not support the theoretical formulation of Taylor and Spence. Thus they are inferring that the study does not support the construct validity of the MAS. In our opinion, a more adequate test of the construct validity of the MAS in their study would require two additional control groups, consisting of more clearly defined high and low *D* groups, where *D* is manipulated experimentally in a nonemotional, positive, approach direction.

Summary

Experimental studies directed at establishing the construct validity of *D*-oriented scales, such as the Taylor MAS, are beset with theoretical and methodological problems that make it difficult to interpret their results. This is particularly true in studies that relate response-inferred, emotionally based *D* to verbal learning. As a partial answer to these problems, the writers contend that studies employing verbal tasks should require a prior demonstration of comparability of the task for extreme groups identified by *D* scale performance. This need is dictated by the emphasis placed on the interaction between *D* level and intratask competition in contemporary *D* theory. The writers further contend that research in this area requires information collected from control groups performing on the same task as the *D* scale groups under high and low *D* conditions that represent clearly defined, experimentally induced motivational states. The information thus provided would serve as a baseline for evaluating the evidence for or against the construct validity of the *D*-oriented scale.

REFERENCES

Bahrick, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, **47**, 170–172.

Child, I. L. Personality. *Ann. Rev. Psychol.*, 1954, **5**, 149–171.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

Hilgard, E. R. Theories of human learning and problems of training. In *Symposium on psychology of learning basic to military training problems*. Panel on Training and Training Devices. Res. & Develpm. Bd., 1953.

Hill, W. F. Comments on Taylor's "Drive theory and manifest anxiety." *Psychol. Bull.*, 1957, **54**, 490–493.

Hughes, J. B., Sprague, J. L., & Bendig, A. W. Anxiety level, response alternation, and performance in serial learning. *J. Psychol.*, 1954, **38**, 421–426.

Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

Jessor, R., & Hammond, K. R. Construct validity and the Taylor Anxiety Scale. *Psychol. Bull.*, 1957, **54**, 161–170.

Katchmar, L. T., Ross, S., & Andrews, T. G. Effects of stress and anxiety on per-

formance of a complex verbal-coding task. *J. exp. Psychol.*, 1958, **55**, 559–564.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.

RAMOND, C. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.

SALTZ, E., & HOEHN, A. J. A test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.

SARASON, I. G. The effect of anxiety and two kinds of failure on serial learning. *J. Pers.*, 1957, **25**, 383–392. (a)

SARASON, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 166–171. (b)

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

SILVERMAN, R. E., & BLITZ, B. Learning and two kinds of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 301–303.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

SPENCE, K. W., FARBER, I. E., MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

TAYLOR, JANET A., & CHAPMAN, J. P. Paired-associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.

TAYLOR, JANET A., &. SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

Received August 8, 1958.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 2, 1959

ON THE CLASSIFICATION OF PROJECTIVE TECHNIQUES[1]

GARDNER LINDZEY
University of Minnesota

Classification is a menial task! It is generally considered to be tedious, unexciting, and an activity, at least for American psychologists, to be delegated to someone else. In spite of this deficiency in allure, and the general tendency of most psychologists to prefer the immediately more rewarding activities of experimental analysis or broad theoretical formulation, there seems little doubt that some emphasis upon classification is important in every branch of psychology. The concern of the present paper is with a relatively humble problem of taxonomy—the classification of projective techniques.

I shall summarize briefly a number of approaches to this problem and suggest a basis for classification that seems to me superior to the various alternatives. Such an enterprise is obviously of interest to those who teach or theorize about projective techniques, for some kind of order must be imposed upon this diverse array of instruments if they are to be discussed efficiently and intelligently. Moreover, a classification that can be agreed upon, and that seems to make psychological sense, should serve some function in research and applied settings, where the interrelation between various projective techniques frequently appears to be a matter of vast importance.

Before turning to a definite proposal in regard to classification, let us consider some of the suggestions that have been made in the past. A number of prior observers have concerned themselves with the ways in which projective techniques can be clustered or grouped, but probably the first and best known of such attempts is contained in the pioneer article by Frank (1939), recommending the use of projective techniques. He suggested that these instruments could be distinguished in terms of whether the responses they elicited were constitutive, interpretative, cathartic, or constructive. The test may be considered *constitutive* if the *S* is required to provide a structure or form for relatively unstructured or ambiguous stimuli, such as finger paints or Rorschach cards. When the *S* is asked to indicate what the meaning of the stimulus is to him, for example, if he is asked to assign meaning to a picture, the test is considered to be *interpretative*. The *cathartic* test involves some deliberate attempt to induce the *S* to express or release emotion in the process of reacting to the stimuli, as in the case of doll play or psychodrama. If the *S* is required to build or organize stimulus materials, such as blocks or toys, in such a manner as to reveal "some of the organizing conceptions of his life" (p. 403) the test is labelled *constructive*. In a subsequent formulation Frank (1948) has added the

[1] This paper is an outgrowth of a monograph on the use of projective techniques in social research initiated by the former Committee on Social Behavior of the Social Science Research Council. Preparation of the article was supported by Research Grant M-1949 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. I am grateful to my colleague Ephraim Rosen for his suggestions and to Arthur Hill for assistance in preparing the tabular material.

category of *refractive* techniques. These are devices that depend upon error or distortion in the *S*'s judgment of some set of stimuli, and are typified by instruments involving tachistoscopic presentation of stimuli related to a given motive or conflict.

Helen Sargent (1945) has considered the problem of classifying projective techniques somewhat more broadly, suggesting that these devices may be grouped in terms of: (*a*) the nature of the materials presented to the *S*s; (*b*) the functional use the *S* makes of the materials; (*c*) the method of presentation used by the examiner; and (*d*) the purpose for which the test is employed. She suggests that under the heading of *materials*, the tests may be distinguished in terms of whether they involve presentation of ink blots, pictures, stories, art media or sounds. The categories proposed under the heading of *functional uses* employed by the *S* are the same as those initially recommended by Frank—constitutive, interpretative, cathartic, and constructive. In distinguishing between tests on the basis of method or *technique of presentation*, Sargent considers variations in both presentation and interpretation. The main distinctions in presentation she discusses have to do with the degree of standardization, or "experimental control," that is imposed upon the examiner and *S*. Differences in interpretive approach that she identifies have to do with an emphasis upon empirical origin, as opposed to theoretical derivation, of the interpretative system. She suggests distinguishing the tests in *purpose* according to whether they are used principally for diagnosis, therapy, or experiment.

In a discussion of the indirect measurement of attitudes, Campbell (1950) suggests three classificatory principles that have implications for projective techniques. First of all, there is the question of whether the device is *disguised* or not, that is, whether the *S* can estimate accurately the intent of the examiner. This dimension might appear to have no utility in the present context, as virtually all projective techniques are assumed to be disguised, but there is actually a moderate degree of variation among projective techniques along this dimension. Second is the question of whether the instrument is *structured* or not. Campbell appears to use this term to refer to both the ambiguity of the stimulus and the amount of freedom permitted the *S* in determining how he will respond. Again these are qualities that are involved in the differentiation of projective techniques from other personality devices. Nevertheless, there is considerable variation between projective techniques along these dimensions and they might well provide the basis for important classificatory distinctions. Third is the distinction between "*voluntary self-description as opposed to* diagnosis based upon *differential performance in an objective task*" (p. 15). Virtually no projective technique can be considered to depend upon "voluntary self description" so that this dimension is of only passing interest in its present form.

Recently, Campbell (1957) has presented a revision of this analysis which is intended to refer more specifically to projective techniques. The new formulation includes three polar dimensions: *voluntary versus objective* (Is the *S* to report something accurately or is he to provide his "own" or "first" response, without regard for correctness?); *indirect versus direct* (Does the *S* know the pur-

pose of the test?); and *free-response versus structured* (Can the S respond much as he chooses or must he select from a limited array of alternatives?). These dimensions are then combined to describe various kinds of psychological tests and Campbell considers examples of the resultant types. He concludes that most projective techniques are voluntary, indirect, and free-response but some projective techniques can be described as voluntary, indirect, and structured; voluntary, direct, and free-response; objective, indirect, and free-response; and objective, indirect, and structured.

Cattell (1951) has suggested that the fundamental process involved in projective tests is not projection but misperception and that these devices should, consequently, be called "misperception techniques." Furthermore, he has indicated that such devices may be divided into four different classes depending upon the form of misperception that operates. First, there is the instrument that depends on *naive misperception*, where the S is unable to recognize the fact that others feel and think differently than he does and, as a result, generalizes his own perceptions to everyone else. Second, there is the test that utilizes the process of *autism*, where the S modifies or distorts his perception in such a manner as to satisfy or reduce his needs and desires. Third are the instruments that involve *press compatibility misperception*, where the S views the environment as existing in such a form as to fit, or make reasonable, his motives and affective states. Fourth are the devices that depend upon *ego defense misperception* where the distortion in perception takes place at the service of unconscious and repressed motives, in a form determined by the various mechanisms of defense.

There are many additional distinctions between projective techniques that can be proposed. For example, we may point to the differences between structural or *formal techniques*, as opposed to "meaning" or *content techniques*. Here the distinction has to do with whether, in interpreting the test, the focus of the examiner is upon the *way* in which the task is performed—the speed and quantity of response, the relative frequency of certain types of words, the tendency to respond to all or to part of the stimulus, etc.—or upon the *meaningful outcome* of the performance. The formal device is concerned with certain quantifiable aspects of the respondent's general pattern of response, and there is little or no interest in the content or meaning of what the respondent is saying or doing. The Rorschach technique is usually considered to be primarily a formal test, although there is considerable evidence for a shift in recent years toward more extensive use of content in interpretation. If the interpretation is focussed upon *what* the individual says or does and its meaning, or the thematic connection between various response elements, the instrument would be classified as a content technique. Illustrative of this type of instrument is the customary use of the Thematic Apperception Test.

Further, we might distinguish between those tests that are *administered individually*, as opposed to those that are capable of *group administration*. Actually, this is a difficult distinction to maintain, for as soon as someone develops an individual technique that seems to possess utility, there are certain to be a number of investigators eagerly seeking to adapt the technique for group administration. Nevertheless, at any

given point in time it is possible to distinguish between tests in terms of how readily they can be adapted to meet the demands of group administration. For example, the sentence completion test can be given in group settings very readily, while doll play or word association techniques are considerably more difficult to administer outside of the individual session.

One might classify projective devices in terms of the *sense modality* involved. For example, there are visual, auditory, and even tactual stimuli employed in tests now in use. An additional distinction between these devices can be made in terms of the degree of *response multiplicity* permitted by the technique. There are a few techniques that require the *S* to choose between a small number of specified alternatives, for example, the Szondi Test; while others permit a theoretically (and almost practically) limitless number of responses, as in the TAT.

Similar to one of the distinctions proposed by Sargent, is the difference between *rational* and *empirical* tests. On the one hand, we have techniques where there is no rationale provided for the fact that a given type of response seems to be associated with a given personality characteristic. Nor is the individual who develops such an instrument concerned with this state of affairs. As long as there is a firm association between a particular type of test response and a given personality variable, he believes that the test may be used in a dependable and useful fashion. The extreme of this approach implies only an interest in the empirical regularity, with no concern for underlying processes or intermediary factors. On the other hand, we have techniques where there is a reasonably careful attempt to provide a prior rational or theoretical basis to justify the use of a particular response element as diagnostic of a particular personality attribute. In practice, it is clear that all devices represent mixtures of these two extremes. The individual who professes to be disinterested in theory and concerned solely with empirical association must make some decision about where to look for his empirical regularity, and here he obviously must drag in some "theory" or prior assumptions. On the other hand, the individual who is interested in a prior rationale, if he is sophisticated, must show considerable curiosity about whether his theoretically predicted relationship is, in fact, sustained in the world of reality and, thus, he introduces the crass empirical criterion. In spite of this overlap, many instruments appear to be more heavily influenced by prior theorizing than others. In general, the Blacky Pictures have developed with a close relationship to explicit theory, while the Rorschach, during at least much of its development, seems to have been treated as an empirical device.

The potential systems of classification we have considered are by no means exhaustive but they serve to illustrate the rich variety that offers itself to the person who surveys this area. How to choose between all of these alternatives? Perhaps an answer to this question can be provided by attempting to classify the classifications. That is, if all the principles of classification can be grouped together, it may be possible to select from among them the avenue that seems most fruitful.

From what has already been said, it is clearly possible to distinguish between six different approaches to the classification of projective techniques. First, there is the distinction

based upon *attributes that inhere in the test material itself.* Here we are concerned with variation in the stimulus material, for example, structured versus unstructured or auditory versus visual. Second, we may classify the tests in terms of the *method by which the technique was devised* or constructed, for example, the distinction between rational and empirical techniques. Third, we may distinguish between these devices on the basis of the *manner in which the test is interpreted*, for example, formal analysis versus content analysis, or "sign" interpretation versus holistic interpretation. Fourth, we might propose a classification that is based upon the *purpose of the test*, for example, the assessment of conflict as opposed to the measurement of motives, or the general description of personality as opposed to the estimation of specified dimensions of personality. Fifth, we might propose a set of categories that are concerned with *differences in the administration of the test*, for example, group technique as opposed to individual technique, or self-administered versus examiner-administered. Sixth, we can distinguish between the instruments on the basis of the *type of response they elicit* from the *S*, for example, story construction as opposed to association.

All of these distinctions have some usefulness and something can be said in favor of each of them as providing the best means for classifying projective techniques. In spite of this, I would argue that the final type of classification, the one based upon differences in type of response, is easily the most important and the one that merits emphasis. The essential consideration here is that this classification seems most likely to be closely related to the underlying psychological processes involved in the various tests, for it is this classification that points to what the *S* is actually doing. In so far as these tests are distinctive, and to be treated as significantly different, it seems likely that the major determinant of this distinctiveness will be the differences in what the *S* is actually engaged in as he responds. It is also worth note that a number of the other types of classification are more or less directly specified by distinctions based upon mode of response, for example, if the technique elicits choice responses, we know a good deal about whether it will emphasize formal or content analysis, whether it is likely to be capable of group administration, and whether it will be structured or not.

Even if we agree that distinctions between projective techniques based upon variation in the type of response elicited from the *S* are most important, there is still the task of arriving at just the proper array of such distinctions. For most purposes it seems sufficient to think in terms of five general types of response. These are: (*a*) association, (*b*) construction, (*c*) completion, (*d*) choice or ordering, and (*e*) expression. Obviously, not every test can be fitted neatly into only one of these categories. There is the usual overlap and ambiguity in the world of reality. However, with very little effort it is possible to classify virtually every projective technique as involving predominantly one of these types of responses. More significant is the fact that when projective techniques are classified on this basis, we find that the instruments brought into the same category have a general congruence and psychological consistency that makes it easily possible to conceive of similar underlying psychological processes.

The reader will note certain similarities between the present categories and those proposed by Frank (1948). However, Frank's categories do not consistently refer to the nature of the *S*'s response, for example, the refractive distinction points to the interpretive process; and, in some cases, the distinction implied by his labels does not seem empirically clear, for example, the distinction between interpretive, constitutive, and constructive is by no means evident. Most important is the fact that his categories do not produce the same clusters of instruments that the present classification generates.

Let us characterize very briefly each of these types of projective techniques and indicate, in an illustrative manner, the individual tests that would be included under each heading. First of all, are the *associative techniques*. Here the *S* is set to respond to some stimulus presented by the examiner with the first word, image or percept that occurs to him. Such devices minimize ideation and emphasize immediacy. The *S* is not to reflect or reason but, rather, to respond with whatever concept or word, however unreasonable, first rises to consciousness, or occurs to him.

These techniques, in certain respects, represent a bridge between experiment and the clinical setting, for in both areas there have been extensive studies of what happens when an individual is asked to respond to some stimulus with the first association that comes to his mind. It was natural that students of the normal, conscious, human mind should use this device as a means of mapping, or laying bare, the structure of mental events. Further, once Freud had devised the method of free association this approach was accepted as an important means of gaining insight into the hidden reaches of the mind. It is not surprising, therefore, that a number of important techniques embodying this mode of response have been developed, the most popular of which are the Word Association Test and the Rorschach test. Also typical of this type of instrument are Stern's Cloud Test and certain auditory projective tests.

Second are the *construction techniques*. Here we find a group of instruments that require the *S* to create or construct a product which is typically an art form such as a story or picture. There is a minimum of restriction placed upon the *S*'s responses and in some cases, such as the blank card of the Thematic Apperception Test, even the original stimulus is under little control by the examiner.

The focus of this type of instrument is upon the outcome, or product, constructed by the *S* and not upon his behavior or style in the process of creating or responding. The *S* is set to provide a product that is meaningful and personally relevant to the eliciting stimuli. The response process may begin with simple association, but the requirements of these tests force the *S* to modify and elaborate the original association, so as to satisfy normative requirements for what constitutes a story or other art form. Unlike the associative techniques, these instruments require the *S* to engage in complex, cognitive activities beyond mere association. Illustrative of these devices are the Thematic Apperception Test, the Blacky Pictures, and the Make-A-Picture-Story Test.

Third, we find the *completion techniques*. These measures provide the *S* with some type of incomplete product and the requirement that he com-

plete it in any manner he wishes. They differ from the associative techniques in that both the stimulus and the response are typically much more complex and thus the response is less immediate. Furthermore, the completed product is usually expected to meet certain external standards of good form or rationality, e.g., there are rules about what constitutes a sentence or a story and they presumably operate to determine the *S*'s completions. When compared to the construction techniques, the responses elicited by these instruments are generally simpler and more restricted. The best known example of this type of instrument is the Sentence Completion Test, but equally typical are the Picture Frustration Study and argument completion and story completion techniques.

Fourth are *choice or ordering techniques*. These instruments resemble the associative measures in the simplicity of the response set provided for the *S*s. Here the respondent merely chooses from a number of alternatives the item or arrangement that fits some specified criterion such as correctness, relevance, attractiveness, or repugnance. In some cases, such as the multiple choice Rorschach and TAT, these devices mirror other techniques except that the *S* is asked not to produce an association or a construction but rather to select from a number of hypothetical responses the one that seems most appropriate to him. The two tests that provide the most effective illustration of this category are the Szondi Test and the Picture Arrangement Test.

Fifth are the *expressive techniques*. As a class, these methods represent a bridge between the diagnostic and therapeutic, for all of them play an active role in current therapeutic practice. It is presumed for these measures that the *S* not only *reveals* himself, but also that he *expresses* himself in such a manner as to influence his personal economy or adjustment. Typically these instruments, as in the case of the constructive techniques, require the *S* to combine or incorporate stimuli into some kind of a novel production. Unlike the constructive techniques, however, there is as much emphasis upon the manner or style in which the product is created, as upon the production itself. In other words, the chief distinction between these measures and constructive devices is the assumption of therapeutic efficacy, and the greater emphasis here upon the style or manner in which the constructive process is carried out. Typical of these instruments are play techniques and drawing and painting techniques, as well as psychodrama and role playing devices.

So much for this simple classification. It is evident that the person who wishes a more complex basis for differentiating projective techniques can readily introduce additional dimensions. Thus, if we return to the six types of classification mentioned earlier, we can easily construct a typology of the sort represented in Table 1. In this table each test is represented by a double plus (++) in the rows that are fully appropriate or descriptive of the test and by a single plus (+) in the rows that are only partially or incompletely appropriate. Thus, for the Rorschach test a double plus (++) in the "formal" row, and a single plus (+) in the "content" row, indicates that both formal and content approaches are used in interpreting the Rorschach, but with greater emphasis placed upon formal variables. The reader should note that we have rated these

TABLE 1

DIMENSIONAL COMPARISON OF PROJECTIVE TECHNIQUES

	Rorschach	Word Assn.	TAT	MAPS	Blacky Pict.	Sent. Complet.	P-F Study	Szondi	Pict. Arrang.	Drawing Painting	Psychodrama
Mode of Response:											
Associative	++	++									
Construction			++	++	++						
Completion						++	++				
Choice or ordering				+	+			++	++		
Expression										++	++
Stimulus Attributes:											
Sensory Mode:											
Visual	++		++	++	++	++	++	++	++	++	+
Auditory		++									++
Tactual										+	
Structured		++	+		+	+	++	+	+		
Unstructured	++			+						++	++
Manner of Interpretation:											
Formal	++	+	+	+	+			++	++	++	
Content	+	++	++	++	++	++	++			+	++
Dimensional	+	++	+		++	++	++	++	++	+	
Holistic	++		++	++		+		+		++	++
Purpose of Test:											
General Personality Description	++		++	++		+		++		++	++
Assessment of Specific Attributes	+	++	+		++	++	++	+			
Identification of Diagnostic Groups	+	+						+	++		
Method of Administration											
Individual	++	++	++	++	++			++		++	++
Group						++	++		++		
Self							+		+		
Restriction of Responses				+	+	+	+	++	++		
Free Responses	++	++	++	++	++					++	++
Method of Construction:											
Rational			+	+	++	+	+	+			
Empirical	+	+				+			++	++	++

tests according to their typical or most highly developed use, in order to maximize differentiation. Therefore, in rating "mode of response" the Rorschach is assigned a double plus (++) for the "associative" row, and no rating for the "choice or ordering" row, in spite of the fact that the test is sometimes given in group forms that involve choice or ordering. It must be admitted that not every one of the judgments registered in this table would meet with unanimous approval among other psychologists, but the majority of the decisions are quite evident and with a little further definition of terms, and specification of standards of judgment, would be made consistently by most trained psychologists.

Thus, we are able to construct a profile for each projective test, representing its classification according to a variety of criteria. Moreover, if we wish, we can readily compute coefficients of similarity or discrepancy to indicate the amount of association between the various instruments on these ratings. Illustrative of such an approach is Table 2 which presents a matrix of deviation coefficients (Osgood's and Suci's D index [1945]) that estimate the degree of association between the instruments as they were rated in Table 1. The reader should note that computation of these indices involved the minor sin of overlooking the fact that our rated dimensions are not orthogonal.

A further step in the analysis is represented in Table 3 where we find illustrative clusters of techniques that seem to be similar to each other in their profiles. All of the tests included in the same cluster are linked by a D index that would place them in the lower quartile (most similar) of the 55 indices presented in Table 2. The clusters are presented in order of decreasing internal homogeneity.

How meaningful are these clusters, and how do they compare with the groupings based on our classification according to type of response? To begin with, it should be noted that three of the clusters (Sentence Completion Test and P-F Study; TAT, MAPS, and Blacky Pictures; Szondi and Picture Arrangement Test) are identical with the clusters that would have been derived from the single criterion of "type of response." Furthermore, it is clear that according to this matrix of indices the Word Association Test is the most individual of all the instruments, for its lowest D index is appreciably higher than those indices linking the tests that we have clustered together. Finally, it turns out that the Rorschach test is singularly difficult to classify. It is clustered with drawing (painting) techniques and psychodrama on the basis of low D indices, but it also shows considerable similarity to the MAPS Test and the TAT, although not with the Blacky Pictures. All in all, the classification that emerges from this somewhat tedious and difficult method of analysis bears a strong resemblance to the simpler classification we have presented, and at the same time it represents a somewhat blurred outcome. There seems little basis at this point for concluding that there is any superiority to such an approach.

To summarize, this paper has pointed to the importance of establishing some consistent basis for classifying projective techniques, and has considered a number of possible approaches to this problem. A classification based upon the mode of response elicited from the *S* was identified as most promising, and it was suggested that projective technique responses can be divided meaning-

TABLE 2

INDICES OF DISCREPANCY BETWEEN PROJECTIVE TECHNIQUES

	Rorschach	Word Assn.	TAT	MAPS	Blacky Pict.	Sent. Complet.	P-F Study	Szondi	Pict. Arrang.	Drawing (Painting)	Psycho-drama
Rorschach											
Word Assn.	5.292										
TAT	4.243	5.477									
MAPS	4.243	6.481	2.449								
Blacky Pict.	5.744	5.000	3.606	4.359							
Sent. Complet.	6.083	5.916	5.099	5.916	5.099						
P-F Study	6.928	5.831	5.831	6.633	5.000	2.236					
Szondi	5.099	6.164	4.899	4.899	4.796	5.385	5.831				
Pict.-Arrang.	6.557	6.708	7.000	6.856	6.324	5.477	5.568	4.582			
Drawing (Painting)	3.464	6.481	4.690	4.472	6.245	6.403	7.348	5.657	6.708		
Psychodrama	4.690	6.000	5.099	4.690	6.708	6.557	7.483	6.782	7.681	3.464	

TABLE 3

PROJECTIVE TECHNIQUES CLUSTERED ACCORDING TO PROFILE SIMILARITY

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Sentence-Completion Test	Thematic Apperception Test	Rorschach	Szondi Test	Word Association Test
P-F Study	Make-A-Picture-Story Test	Drawing and Painting Techniques	Picture Arrangement Test	
	Blacky Pictures	Psychodrama		

fully according to whether they involve: association, construction, completion, choice (ordering), or expression. Moreover, when this classification was compared with a more complex (multidimensional) taxonomy there seemed to be little basis for preferring the more cumbersome method of classification.

REFERENCES

CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15–38.

CAMPBELL, D. T. A typology of tests, projective and otherwise. *J. consult. Psychol.*, 1957, **21**, 207–210.

CATTELL, R. B. Principles of design in "projective" or misperception tests of personality. In H. H. Anderson and Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 55–98.

FRANK, L. K. Projective methods for the study of personality. *J. Psychol.*, 1939, **8**, 389–413.

FRANK, L. K. *Projective methods*. Springfield, Ill.: Thomas, 1948.

OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

SARGENT, HELEN. Projective methods; their origins, theory, and application in personality research. *Psychol. Bull.*, 1945, **42**, 257–293.

Received August 22, 1958.

MICROFORM

Journals of the American Psychological Association
available on MICROFILM or MICROCARD

Psychological Review

American Psychologist

Psychological Bulletin

Psychological Abstracts

Contemporary Psychology

Psychological Monographs

Journal of Applied Psychology

Journal of Consulting Psychology

Journal of Educational Psychology

Journal of Experimental Psychology

Journal of Abnormal and Social Psychology

Journal of Comparative and Physiological Psychology

Available only in volume units; no single issues.

For MICROFILM, order from:

UNIVERSITY
MICROFILMS, INC.
313 North First Street
Ann Arbor,
Michigan

For MICROCARD, order from:

J. S. CANNER &
COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN, U.S.A.